

1. [Preface](#)
2. Welcome to Economics!
 1. [Introduction](#)
 2. [What Is Economics, and Why Is It Important?](#)
 3. [Microeconomics and Macroeconomics](#)
 4. [How Economists Use Theories and Models to Understand Economic Issues](#)
 5. [How Economies Can Be Organized: An Overview of Economic Systems](#)
3. Choice in a World of Scarcity
 1. [Introduction to Choice in a World of Scarcity](#)
 2. [How Individuals Make Choices Based on Their Budget Constraint](#)
 3. [The Production Possibilities Frontier and Social Choices](#)
 4. [Confronting Objections to the Economic Approach](#)
4. Demand and Supply
 1. [Introduction to Demand and Supply](#)
 2. [Demand, Supply, and Equilibrium in Markets for Goods and Services](#)
 3. [Shifts in Demand and Supply for Goods and Services](#)
 4. [Price Ceilings and Price Floors](#)
5. Elasticity
 1. [Introduction to Elasticity](#)
 2. [Price Elasticity of Demand and Price Elasticity of Supply](#)
 3. [Elasticity and Pricing](#)
6. The Macroeconomic Perspective
 1. [Introduction to the Macroeconomic Perspective](#)
 2. [Measuring the Size of the Economy: Gross Domestic Product](#)
 3. [Adjusting Nominal Values to Real Values](#)
 4. [Tracking Real GDP over Time](#)

5. [Comparing GDP among Countries](#)

7. Unemployment

1. [Introduction to Unemployment](#)
2. [How the Unemployment Rate is Defined and Computed](#)
3. [Patterns of Unemployment](#)
4. [What Causes Changes in Unemployment over the Short Run](#)
5. [What Causes Changes in Unemployment over the Long Run](#)

8. Inflation

1. [Introduction to Inflation](#)
2. [Tracking Inflation](#)
3. [How Changes in the Cost of Living are Measured](#)
4. [How the U.S. and Other Countries Experience Inflation](#)
5. [The Confusion Over Inflation](#)
6. [Indexing and Its Limitations](#)

9. Government Budgets and Fiscal Policy

1. [Introduction to Government Budgets and Fiscal Policy](#)
2. [Government Spending](#)
3. [Taxation](#)
4. [Federal Deficits and the National Debt](#)
5. [Using Fiscal Policy to Fight Recession, Unemployment, and Inflation](#)
6. [The Multiplier Effect](#)
7. [Automatic Stabilizers](#)
8. [Practical Problems with Discretionary Fiscal Policy](#)
9. [The Question of a Balanced Budget](#)

10. Money and Banking

1. [Introduction to Money and Banking](#)
2. [Defining Money by Its Functions](#)
3. [Measuring Money: Currency, M1, and M2](#)
4. [The Role of Banks](#)

5. [How Banks Create Money](#)

11. Monetary Policy and Bank Regulation

1. [Introduction to Monetary Policy and Bank Regulation](#)
2. [The Federal Reserve Banking System and Central Banks](#)
3. [Bank Regulation](#)
4. [How a Central Bank Executes Monetary Policy](#)
5. [Monetary Policy and Economic Outcomes](#)
6. [Pitfalls for Monetary Policy](#)

12. The Keynesian Perspective

1. [Introduction to the Keynesian Perspective](#)
2. [Aggregate Demand in Keynesian Analysis](#)
3. [The Building Blocks of Keynesian Analysis](#)
4. [The Phillips Curve](#)
5. [The Keynesian Perspective on Market Forces](#)

13. The Neoclassical Perspective

1. [Introduction to the Neoclassical Perspective](#)
2. [The Building Blocks of Neoclassical Analysis](#)
3. [The Policy Implications of the Neoclassical Perspective](#)
4. [Balancing Keynesian and Neoclassical Models](#)

14. Environmental Protection and Negative Externalities

1. [Introduction to Environmental Protection and Negative Externalities](#)
2. [The Economics of Pollution](#)
3. [Command-and-Control Regulation](#)
4. [Market-Oriented Environmental Tools](#)
5. [The Benefits and Costs of U.S. Environmental Laws](#)

15. Positive Externalities and Public Goods

1. [Introduction to Positive Externalities and Public Goods](#)
2. [Why the Private Sector Under Invests in Innovation](#)
3. [Why the Private Sector Under Invests in Innovation](#)

16. Information, Risk, and Insurance

1. [Introduction to Information, Risk, and Insurance](#)

2. [The Problem of Imperfect Information and Asymmetric Information](#)
 3. [Insurance and Imperfect Information](#)
17. Perfect Competition
 1. [Perfect Competition and Why It Matters](#)
 2. [Perfect Competition and Why It Matters](#)
 3. [Efficiency in Perfectly Competitive Markets](#)
18. Monopoly
 1. [Introduction to a Monopoly](#)
 2. [How Monopolies Form: Barriers to Entry](#)
 3. [How a Profit-Maximizing Monopoly Chooses Output and Price](#)
19. Monopolistic Competition and Oligopoly
 1. [Introduction to Monopolistic Competition and Oligopoly](#)
 2. [Monopolistic Competition](#)
 3. [Oligopoly](#)
20. Monopoly and Antitrust Policy
 1. [Introduction to Monopoly and Antitrust Policy](#)
 2. [Corporate Mergers](#)
 3. [Regulating Anticompetitive Behavior](#)
 4. [Regulating Natural Monopolies](#)
 5. [The Great Deregulation Experiment](#)
21. The International Trade and Capital Flows
 1. [Introduction to the International Trade and Capital Flows](#)
 2. [Measuring Trade Balances](#)
 3. [Trade Balances in Historical and International Context](#)
 4. [Trade Balances and Flows of Financial Capital](#)
 5. [The Pros and Cons of Trade Deficits and Surpluses](#)
 6. [The Difference between Level of Trade and the Trade Balance](#)
22. Exchange Rates and International Capital Flows

1. [Introduction to Exchange Rates and International Capital Flows](#)
2. [How the Foreign Exchange Market Works](#)
3. [Demand and Supply Shifts in Foreign Exchange Markets](#)
4. [Macroeconomic Effects of Exchange Rates](#)
5. [Exchange Rate Policies](#)

23. International Trade

1. [Introduction to International Trade](#)
2. [Absolute and Comparative Advantage](#)
3. [What Happens When a Country Has an Absolute Advantage in All Goods](#)
4. [Intra-industry Trade between Similar Economies](#)
5. [The Benefits of Reducing Barriers to International Trade](#)

Preface

Principles of Economics is designed for a two-semester principles of economics sequence. It is traditional in coverage, including introductory economics content, microeconomics, macroeconomics and international economics. At the same time, the book includes a number of innovative and interactive features designed to enhance student learning. Instructors can also customize the book, adapting it to the approach that works best in their classroom.

Editor of the FVTC Edition

Rick Reid, Fox Valley Technical College

Rick Reid is the editor of Principles of Economics-FVTC edition. Rick has been teaching Principles of Economics courses at Fox Valley Technical College and Lakeland University since 2010, and is currently the committee chair for the Economics section at Fox Valley Tech. He earned his Bachelors of Arts degree in Economics from the University of Illinois, and his MBA from Minnesota State University. Prior to teaching, Rick held a variety of positions within the pharmaceutical and retail industries. Rick Reid lives in the Fox Valley of Wisconsin, with Jean, his loving wife of 35 years. Without Jean's support, Rick wouldn't have been able to complete his graduate degree, much less this current project.

Thank You From the Editor

A special thanks to all those at Fox Valley Technical College who were involved in the production of Principles of Economics-FVTC edition. In particular, to my good friend and valued colleague, Mia Gauthier, who provided a 'second set of eyes' on edits, corrections, decisions and sage advice that contributed to this publication. An additional note of thanks to all of the folks at Openstax, who provided ongoing support before and during the editing process, answering questions and clarifying processes and parameters of publication. Without the generous support of funding organizations that support the Openstax initiative, this type of resource

would not be possible. It is our sincere hope that this published work holds value for all of the students and instructors who use it.

Dedication

My work on the Principles of Economics-FVTC edition is an homage to my parents, John and Sara Reid. Impacted by both the Great Depression and World War II, neither of them had the means to pursue a college degree, so they spent most of their working lives ensuring that all three of their children had both the opportunity and the expectation of earning a degree. All three of us did and so have all of our children. It is my sincere hope that ready access to a high quality textbook such as this helps others in their pursuit of a formal education and the benefits which that confers.

About OpenStax

OpenStax is a non-profit organization committed to improving student access to quality learning materials. Our free textbooks go through a rigorous editorial publishing process. Our texts are developed and peer-reviewed by educators to ensure they are readable, accurate, and meet the scope and sequence requirements of today's college courses. Unlike traditional textbooks, OpenStax resources live online and are owned by the community of educators using them. Through our partnerships with companies and foundations committed to reducing costs for students, OpenStax is working to improve access to higher education for all. OpenStax is an initiative of Rice University and is made possible through the generous support of several philanthropic foundations.

About OpenStax's Resources

Welcome to *Principles of Economics*, an OpenStax resource. This textbook has been created with several goals in mind: accessibility, customization, and student engagement—all while encouraging students toward high levels of academic scholarship. Instructors and students alike will find that this textbook offers a strong foundation in economics in an accessible format.

OpenStax resources provide quality academic instruction. Three key features set our materials apart from others: they can be customized by instructors for each class, they are a "living" resource that grows online through contributions from science educators, and they are available free or for minimal cost.

Customization

OpenStax learning resources are designed to be customized for each course. Our textbooks provide a solid foundation on which instructors can build, and our resources are conceived and written with flexibility in mind. Instructors can select the sections most relevant to their curricula and create a textbook that speaks directly to the needs of their classes and student body. Teachers are encouraged to expand on existing examples by adding unique context via geographically localized applications and topical connections.

Principles of Economics can be easily customized using our online platform (<http://cnx.org/content/col11613/>). Simply select the content most relevant to your current semester and create a textbook that speaks directly to the needs of your class. *Principles of Economics* is organized as a collection of sections that can be rearranged, modified, and enhanced through localized examples or to incorporate a specific theme of your course. This customization feature will ensure that your textbook truly reflects the goals of your course. *Principles of Economics* is also available in two volumes, one covering microeconomics and one covering macroeconomics principles.

Curation

To broaden access and encourage community curation, *Principles of Economics* is “open source” licensed under a Creative Commons Attribution (CC-BY) license. The economics community is invited to submit examples, emerging research, and other feedback to enhance and

strengthen the material and keep it current and relevant for today's students. You can submit your suggestions to info@openstaxcollege.org.

Cost

Our textbooks are available for free online, and in low-cost print and e-book editions.

About *Principles of Economics*

This version of *Principles of Economics* (Principles of Economics-FVTC), has been edited to serve a one semester Principles of Economics class. The book on which it is based was designed for a two-semester principles of economics sequence. The text has been developed to meet the scope and sequence of most introductory courses. At the same time, the book includes a number of innovative features designed to enhance student learning.

Coverage and Scope

To develop *Principles of Economics*, we acquired the rights to Timothy Taylor's second edition of *Principles of Economics* and solicited ideas from economics instructors at all levels of higher education, from community colleges to Ph.D.-granting universities. They told us about their courses, students, challenges, resources, and how a textbook can best meet the needs of both instructors and students.

The result is a book that covers the breadth of economics topics and also provides the necessary depth to ensure the course is manageable for instructors and students alike. And to make it more applied, we have incorporated many current topics. We hope students will be interested to know just how far-reaching the recent recession was (and still is), for example, and why there is so much controversy even among economists over the Affordable Care Act (Obamacare). The Keystone Pipeline, Occupy Wall Street, minimum wage debates, and the appointment of the United

States' first female Federal Reserve chair, Janet Yellen, are just a few of the other important topics covered.

The pedagogical choices, chapter arrangements, and learning objective fulfillment were developed and vetted with feedback from educators dedicated to the project. They thoroughly read the material and offered critical and detailed commentary. The outcome is a balanced approach to micro and macro economics, to both Keynesian and classical views, and to the theory and application of economics concepts. New 2015 data are incorporated for topics that range from average U.S. household consumption in Chapter 2 to the total value of all home equity in Chapter 17. Current events are treated in a politically-balanced way as well.

Chapters in the original version:

- Chapter 1 Welcome to Economics!
- Chapter 2 Choice in a World of Scarcity
- Chapter 3 Demand and Supply
- Chapter 4 Labor and Financial Markets
- Chapter 5 Elasticity
- Chapter 6 Consumer Choices
- Chapter 7 Cost and Industry Structure
- Chapter 8 Perfect Competition
- Chapter 9 Monopoly
- Chapter 10 Monopolistic Competition and Oligopoly
- Chapter 11 Monopoly and Antitrust Policy
- Chapter 12 Environmental Protection and Negative Externalities
- Chapter 13 Positive Externalities and Public Goods
- Chapter 14 Poverty and Economic Inequality
- Chapter 15 Issues in Labor Markets: Unions, Discrimination, Immigration
- Chapter 16 Information, Risk, and Insurance
- Chapter 17 Financial Markets
- Chapter 18 Public Economy
- Chapter 19 The Macroeconomic Perspective
- Chapter 20 Economic Growth
- Chapter 21 Unemployment
- Chapter 22 Inflation
- Chapter 23 The International Trade and Capital Flows

Chapter 24 The Aggregate Demand/Aggregate Supply Model
Chapter 25 The Keynesian Perspective
Chapter 26 The Neoclassical Perspective
Chapter 27 Money and Banking
Chapter 28 Monetary Policy and Bank Regulation
Chapter 29 Exchange Rates and International Capital Flows
Chapter 30 Government Budgets and Fiscal Policy
Chapter 31 The Impacts of Government Borrowing
Chapter 32 Macroeconomic Policy Around the World
Chapter 33 International Trade
Chapter 34 Globalization and Protectionism

Appendix A The Use of Mathematics in Principles of Economics
Appendix B Indifference Curves
Appendix C Present Discounted Value
Appendix D The Expenditure-Output Model

Pedagogical Foundation

Throughout the OpenStax version of *Principles of Economics*, you will find new features that engage the students in economic inquiry by taking selected topics a step further. Our features include:

Bring It Home: This added feature is a brief case study, specific to each chapter, which connects the chapter's main topic to the real world. It is broken up into two parts: the first at the beginning of the chapter (in the Intro module) and the second at chapter's end, when students have learned what's necessary to understand the case and "bring home" the chapter's core concepts.

Work It Out: This added feature asks students to work through a generally analytical or computational problem, and guides them step-by-step to find out how its solution is derived.

Clear It Up: This boxed feature, which includes pre-existing features from Taylor's text, addresses common student misconceptions about the content. Clear It Ups are usually deeper explanations of something

in the main body of the text. Each CIU starts with a question. The rest of the feature explains the answer.

Link It Up: This added feature is a very brief introduction to a website that is pertinent to students' understanding and enjoyment of the topic at hand.

Questions for Each Level of Learning

The OpenStax version of *Principles of Economics* further expands on Taylor's original end of chapter materials by offering four types of end-of-module questions for students.

Self-Checks: Are analytical self-assessment questions that appear at the end of each module. They “click-to-reveal” an answer in the web view so students can check their understanding before moving on to the next module. Self-Check questions are not simple look-up questions. They push the student to think a bit beyond what is said in the text. Self-Check questions are designed for formative (rather than summative) assessment. The questions and answers are explained so that students feel like they are being walked through the problem.

Review Questions: Have been retained from Taylor's version, and are simple recall questions from the chapter and are in open-response format (not multiple choice or true/false). The answers can be looked up in the text.

Critical Thinking Questions: Are new higher-level, conceptual questions that ask students to *demonstrate their understanding by applying* what they have learned in different contexts. They ask for outside-the-box thinking, for *reasoning* about the concepts. They push the student to places they wouldn't have thought of going themselves.

Problems: Are exercises that give students additional practice working with the analytic and computational concepts in the module.

Updated Art

Principles of Economics includes an updated art program to better inform today's student, providing the latest data on covered topics.

About Our Team

Senior Contributing Author

Timothy Taylor, Macalester College

Timothy Taylor has been writing and teaching about economics for 30 years, and is the Managing Editor of the *Journal of Economic Perspectives*, a post he's held since 1986. He has been a lecturer for The Teaching Company, the University of Minnesota, and the Hubert H. Humphrey Institute of Public Affairs, where students voted him Teacher of the Year in 1997. His writings include numerous pieces for journals such as the *Milken Institute Review* and *The Public Interest*, and he has been an editor on many projects, most notably for the Brookings Institution and the World Bank, where he was Chief Outside Editor for the *World Development Report 1999/2000, Entering the 21st Century: The Changing Development Landscape*. He also blogs four to five times per week at <http://conversableeconomist.blogspot.com>. Timothy Taylor lives near Minneapolis with his wife Kimberley and their three children.

Steven A. Greenlaw, University of Mary Washington

Steven Greenlaw has been teaching principles of economics for more than 30 years. In 1999, he received the Grellet C. Simpson Award for Excellence in Undergraduate Teaching at the University of Mary Washington. He is the author of *Doing Economics: A Guide to Doing and Understanding Economic Research*, as well as a variety of articles on economics pedagogy and instructional technology, published in the *Journal of Economic Education*, the *International Review of Economic Education*, and other outlets. He wrote the module on Quantitative Writing for *Starting Point: Teaching and Learning Economics*, the web portal on best practices in teaching economics. Steven Greenlaw lives in Alexandria, Virginia with his wife Kathy and their three children.

Contributing Authors

Eric Dodge, Hanover College
Cynthia Gamez, University of Texas at El Paso
Andres Jauregui, Columbus State University
Diane Keenan, Cerritos College
Dan MacDonald, California State University San Bernardino
Amyaz Moledina, The College of Wooster
Craig Richardson, Winston-Salem State University
David Shapiro, Pennsylvania State University
Ralph Sonenshine, American University

Expert Reviewers

Bryan Aguiar, Northwest Arkansas Community College
Basil Al Hashimi, Mesa Community College
Emil Berendt, Mount St. Mary's University
Zena Buser, Adams State University
Douglas Campbell, The University of Memphis
Sanjukta Chaudhuri, University of Wisconsin - Eau Claire
Xueyu Cheng, Alabama State University
Robert Cunningham, Alma College
Rosa Lea Danielson, College of DuPage
Steven Deloach, Elon University
Debbie Evercloud, University of Colorado Denver
Sal Figueras, Hudson County Community College
Reza Ghorashi, Richard Stockton College of New Jersey
Robert Gillette, University of Kentucky
Shaomin Huang, Lewis-Clark State College
George Jones, University of Wisconsin-Rock County
Charles Kroncke, College of Mount St. Joseph
Teresa Laughlin, Palomar Community College
Carlos Liard-Muriente, Central Connecticut State University
Heather Luea, Kansas State University
Steven Lugauer, University of Notre Dame
William Mosher, Nashua Community College

Michael Netta, Hudson County Community College
Nick Noble, Miami University
Joe Nowakowski, Muskingum University
Shawn Osell, University of Wisconsin, Superior
Mark Owens, Middle Tennessee State University
Sonia Pereira, Barnard College
Brian Peterson, Central College
Jennifer Platania, Elon University
Robert Rycroft, University of Mary Washington
Adrienne Sachse, Florida State College at Jacksonville
Hans Schumann, Texas AM
Gina Shamshak, Goucher College
Chris Warburton, John Jay College of Criminal Justice, CUNY
Mark Witte, Northwestern
Chiou-nan Yeh, Alabama State University

Ancillaries

OpenStax projects offer an array of ancillaries for students and instructors. Please visit <http://openstaxcollege.org> and view the learning resources for this title.

Introduction
class="introduction"
Do You Use Facebook?

Economics is
greatly
impacted by
how well
information
travels through
society. Today,
social media
giants Twitter,
Facebook, and
Instagram are
major forces
on the
information
super highway.
(Credit: Johan
Larsson/Flickr
)

**Note:****Decisions ... Decisions in the Social Media Age**

To post or not to post? Every day we are faced with a myriad of decisions, from what to have for breakfast, to which route to take to class, to the more complex—“Should I double major and add possibly another semester of study to my education?” Our response to these choices depends on the information we have available at any given moment; information economists call “imperfect” because we rarely have all the data we need to make perfect decisions. Despite the lack of perfect information, we still make hundreds of decisions a day.

And now, we have another avenue in which to gather information—social media. Outlets like Facebook and Twitter are altering the process by which we make choices, how we spend our time, which movies we see, which products we buy, and more. How many of you chose a university without checking out its Facebook page or Twitter stream first for information and feedback?

As you will see in this course, what happens in economics is affected by how well and how fast information is disseminated through a society, such as how quickly information travels through Facebook. “Economists love nothing better than when deep and liquid markets operate under conditions of perfect information,” says Jessica Irvine, National Economics Editor for News Corp Australia.

This leads us to the topic of this chapter, an introduction to the world of making decisions, processing information, and understanding behavior in markets—the world of economics. Each chapter in this book will start with a discussion about current (or sometimes past) events and revisit it at chapter’s end—to “bring home” the concepts in play.

Note:**Introduction**

In this chapter, you will learn about:

- What Is Economics, and Why Is It Important?
- Microeconomics and Macroeconomics
- How Economists Use Theories and Models to Understand Economic Issues
- How Economies Can Be Organized: An Overview of Economic Systems

What is economics and why should you spend your time learning it? After all, there are other disciplines you could be studying, and other ways you could be spending your time. As the Bring it Home feature just mentioned, making choices is at the heart of what economists study, and your decision to take this course is as much an economic decision as anything else.

Economics is probably not what you think. It is not primarily about money or finance. It is not primarily about business. It is not mathematics. What is it then? It is both a subject area and a way of viewing the world.

What Is Economics, and Why Is It Important?

By the end of this section, you will be able to:

- Discuss the importance of studying economics
- Explain the relationship between production and division of labor
- Evaluate the significance of scarcity

Economics is the study of how humans make decisions in the face of scarcity. These can be individual decisions, family decisions, business decisions or societal decisions. If you look around carefully, you will see that scarcity is a fact of life. **Scarcity** means that human wants for goods, services and resources exceed what is available. Resources, such as labor, tools, land, and raw materials are necessary to produce the goods and services we want but they exist in limited supply. Of course, the ultimate scarce resource is time- everyone, rich or poor, has just 24 hours in the day to try to acquire the goods they want. At any point in time, there is only a finite amount of resources available.

Think about it this way: In 2015 the labor force in the United States contained over 158.6 million workers, according to the U.S. Bureau of Labor Statistics. Similarly, the total area of the United States is 3,794,101 square miles. These are large numbers for such crucial resources, however, they are limited. Because these resources are limited, so are the numbers of goods and services we produce with them. Combine this with the fact that human wants seem to be virtually infinite, and you can see why scarcity is a problem.

Scarcity of Resources



Homeless people are a stark reminder that scarcity of resources is real. (Credit: “daveynin”/Flickr Creative Commons)

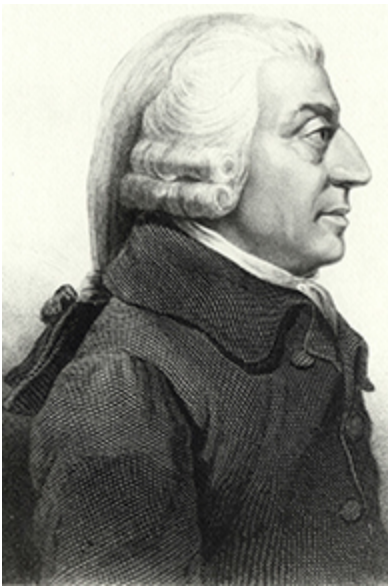
If you still do not believe that scarcity is a problem, consider the following: Does everyone need food to eat? Does everyone need a decent place to live? Does everyone have access to healthcare? In every country in the world, there are people who are hungry, homeless (for example, those who call park benches their beds, as shown in [\[link\]](#)), and in need of healthcare, just to focus on a few critical goods and services. Why is this the case? It is because of scarcity. Let’s delve into the concept of scarcity a little deeper, because it is crucial to understanding economics.

The Problem of Scarcity

Think about all the things you consume: food, shelter, clothing, transportation, healthcare, and entertainment. How do you acquire those items? You do not produce them yourself. You buy them. How do you afford the things you buy? You work for pay. Or if you do not, someone else does on your behalf. Yet most of us never have enough to buy all the things we want. This is because of scarcity. So how do we solve it?

Every society, at every level, must make choices about how to use its resources. Families must decide whether to spend their money on a new car or a fancy vacation. Towns must choose whether to put more of the budget into police and fire protection or into the school system. Nations must decide whether to devote more funds to national defense or to protecting the environment. In most cases, there just isn't enough money in the budget to do everything. So why do we not each just produce all of the things we consume? The simple answer is most of us do not know how, but that is not the main reason. (When you study economics, you will discover that the obvious choice is not always the right answer—or at least the complete answer. Studying economics teaches you to think in a different of way.) Think back to pioneer days, when individuals knew how to do so much more than we do today, from building their homes, to growing their crops, to hunting for food, to repairing their equipment. Most of us do not know how to do all—or any—of those things. It is not because we could not learn. Rather, we do not have to. The reason why is something called *the division and specialization of labor*, a production innovation first put forth by Adam Smith, [\[link\]](#), in his book, *The Wealth of Nations*.

Adam Smith



Adam Smith
introduced the idea
of dividing labor
into discrete tasks.

(Credit:
Wikimedia
Commons)

The Division of and Specialization of Labor

The formal study of economics began when Adam Smith (1723–1790) published his famous book *The Wealth of Nations* in 1776. Many authors had written on economics in the centuries before Smith, but he was the first to address the subject in a comprehensive way. In the first chapter, Smith introduces the **division of labor**, which means that the way a good or service is produced is divided into a number of tasks that are performed by different workers, instead of all the tasks being done by the same person.

To illustrate the division of labor, Smith counted how many tasks went into making a pin: drawing out a piece of wire, cutting it to the right length, straightening it, putting a head on one end and a point on the other, and packaging pins for sale, to name just a few. Smith counted 18 distinct tasks that were often done by different people—all for a pin, believe it or not!

Modern businesses divide tasks as well. Even a relatively simple business like a restaurant divides up the task of serving meals into a range of jobs like top chef, sous chefs, less-skilled kitchen help, servers to wait on the tables, a greeter at the door, janitors to clean up, and a business manager to handle paychecks and bills—not to mention the economic connections a restaurant has with suppliers of food, furniture, kitchen equipment, and the building where it is located. A complex business like a large manufacturing factory, such as the shoe factory shown in [\[link\]](#), or a hospital can have hundreds of job classifications.

Division of Labor



Workers on an assembly line are an example of the divisions of labor. (Credit: Nina Hale/Flickr Creative Commons)

Why the Division of Labor Increases Production

When the tasks involved with producing a good or service are divided and subdivided, workers and businesses can produce a greater quantity of output. In his observations of pin factories, Smith observed that one worker alone might make 20 pins in a day, but that a small business of 10 workers (some of whom would need to do two or three of the 18 tasks involved with pin-making), could make 48,000 pins in a day. How can a group of workers, each specializing in certain tasks, produce so much more than the same number of workers who try to produce the entire good or service by themselves? Smith offered three reasons.

First, **specialization** in a particular small job allows workers to focus on the parts of the production process where they have an advantage. (In later chapters, we will develop this idea by discussing comparative advantage.) People have different skills, talents, and interests, so they will be better at some jobs than at others. The particular advantages may be based on educational choices, which are in turn shaped by interests and talents. Only

those with medical degrees qualify to become doctors, for instance. For some goods, specialization will be affected by geography—it is easier to be a wheat farmer in North Dakota than in Florida, but easier to run a tourist hotel in Florida than in North Dakota. If you live in or near a big city, it is easier to attract enough customers to operate a successful dry cleaning business or movie theater than if you live in a sparsely populated rural area. Whatever the reason, if people specialize in the production of what they do best, they will be more productive than if they produce a combination of things, some of which they are good at and some of which they are not.

Second, workers who specialize in certain tasks often learn to produce more quickly and with higher quality. This pattern holds true for many workers, including assembly line laborers who build cars, stylists who cut hair, and doctors who perform heart surgery. In fact, specialized workers often know their jobs well enough to suggest innovative ways to do their work faster and better.

A similar pattern often operates within businesses. In many cases, a business that focuses on one or a few products (sometimes called its “core competency”) is more successful than firms that try to make a wide range of products.

Third, specialization allows businesses to take advantage of **economies of scale**, which means that for many goods, as the level of production increases, the average cost of producing each individual unit declines. For example, if a factory produces only 100 cars per year, each car will be quite expensive to make on average. However, if a factory produces 50,000 cars each year, then it can set up an assembly line with huge machines and workers performing specialized tasks, and the average cost of production per car will be lower. The ultimate result of workers who can focus on their preferences and talents, learn to do their specialized jobs better, and work in larger organizations is that society as a whole can produce and consume far more than if each person tried to produce all of their own goods and services. The division and specialization of labor has been a force against the problem of scarcity.

Trade and Markets

Specialization only makes sense, though, if workers can use the pay they receive for doing their jobs to purchase the other goods and services that they need. In short, specialization requires trade.

You do not have to know anything about electronics or sound systems to play music—you just buy an iPod or MP3 player, download the music and listen. You do not have to know anything about artificial fibers or the construction of sewing machines if you need a jacket—you just buy the jacket and wear it. You do not need to know anything about internal combustion engines to operate a car—you just get in and drive. Instead of trying to acquire all the knowledge and skills involved in producing all of the goods and services that you wish to consume, the market allows you to learn a specialized set of skills and then use the pay you receive to buy the goods and services you need or want. This is how our modern society has evolved into a strong economy.

Why Study Economics?

Now that we have gotten an overview on what economics studies, let's quickly discuss why you are right to study it. Economics is not primarily a collection of facts to be memorized, though there are plenty of important concepts to be learned. Instead, economics is better thought of as a collection of questions to be answered or puzzles to be worked out. Most important, economics provides the tools to work out those puzzles. If you have yet to be bitten by the economics “bug,” there are other reasons why you should study economics.

- Virtually every major problem facing the world today, from global warming, to world poverty, to the conflicts in Syria, Afghanistan, and Somalia, has an economic dimension. If you are going to be part of solving those problems, you need to be able to understand them. Economics is crucial.
- It is hard to overstate the importance of economics to good citizenship. You need to be able to vote intelligently on budgets, regulations, and laws in general. When the U.S. government came close to a standstill at the end of 2012 due to the “fiscal cliff,” what were the issues involved? Did you know?

- A basic understanding of economics makes you a well-rounded thinker. When you read articles about economic issues, you will understand and be able to evaluate the writer's argument. When you hear classmates, co-workers, or political candidates talking about economics, you will be able to distinguish between common sense and nonsense. You will find new ways of thinking about current events and about personal and business decisions, as well as current events and politics.

The study of economics does not dictate the answers, but it can illuminate the different choices.

Key Concepts and Summary

Economics seeks to solve the problem of scarcity, which is when human wants for goods and services exceed the available supply. A modern economy displays a division of labor, in which people earn income by specializing in what they produce and then use that income to purchase the products they need or want. The division of labor allows individuals and firms to specialize and to produce more for several reasons: a) It allows the agents to focus on areas of advantage due to natural factors and skill levels; b) It encourages the agents to learn and invent; c) It allows agents to take advantage of economies of scale. Division and specialization of labor only work when individuals can purchase what they do not produce in markets. Learning about economics helps you understand the major problems facing the world today, prepares you to be a good citizen, and helps you become a well-rounded thinker.

Self-Check Questions

Exercise:

Problem: What is scarcity? Can you think of two causes of scarcity?

Solution:

Scarcity means human wants for goods and services exceed the available supply. Supply is limited because resources are limited. Demand, however, is virtually unlimited. Whatever the supply, it seems human nature to want more.

Exercise:

Problem:

Residents of the town of Smithfield like to consume hams, but each ham requires 10 people to produce it and takes a month. If the town has a total of 100 people, what is the maximum amount of ham the residents can consume in a month?

Solution:

$100 \text{ people} / 10 \text{ people per ham} = \text{a maximum of 10 hams per month}$ if all residents produce ham. Since consumption is limited by production, the maximum number of hams residents could consume per month is 10.

Exercise:

Problem:

A consultant works for \$200 per hour. She likes to eat vegetables, but is not very good at growing them. Why does it make more economic sense for her to spend her time at the consulting job and shop for her vegetables?

Solution:

She is very productive at her consulting job, but not very productive growing vegetables. Time spent consulting would produce far more income than it what she could save growing her vegetables using the same amount of time. So on purely economic grounds, it makes more sense for her to maximize her income by applying her labor to what she does best (i.e. specialization of labor).

Exercise:

Problem:

A computer systems engineer could paint his house, but it makes more sense for him to hire a painter to do it. Explain why.

Solution:

The engineer is better at computer science than at painting. Thus, his time is better spent working for pay at his job and paying a painter to paint his house. Of course, this assumes he does not paint his house for fun!

Review Questions**Exercise:****Problem:**

Give the three reasons that explain why the division of labor increases an economy's level of production.

Exercise:

Problem: What are three reasons to study economics?

Critical Thinking Questions**Exercise:****Problem:**

Suppose you have a team of two workers: one is a baker and one is a chef. Explain why the kitchen can produce more meals in a given period of time if each worker specializes in what they do best than if each worker tries to do everything from appetizer to dessert.

Exercise:

Problem: Why would division of labor without trade not work?

Exercise:

Problem:

Can you think of any examples of *free* goods, that is, goods or services that are not scarce?

References

Bureau of Labor Statistics, U.S. Department of Labor. 2015. "The Employment Situation—February 2015." Accessed March 27, 2015.
<http://www.bls.gov/news.release/pdf/empstat.pdf>.

Williamson, Lisa. "US Labor Market in 2012." *Bureau of Labor Statistics*. Accessed December 1, 2013.
<http://www.bls.gov/opub/mlr/2013/03/art1full.pdf>.

Glossary

division of labor

the way in which the work required to produce a good or service is divided into tasks performed by different workers

economics

the study of how humans make choices under conditions of scarcity

economies of scale

when the average cost of producing each individual unit declines as total output increases

scarcity

when human wants for goods and services exceed the available supply

specialization

when workers or firms focus on particular tasks for which they are well-suited within the overall production process

Microeconomics and Macroeconomics

By the end of this section, you will be able to:

- Describe microeconomics
- Describe macroeconomics
- Contrast monetary policy and fiscal policy

Economics is concerned with the well-being of *all* people, including those with jobs and those without jobs, as well as those with high incomes and those with low incomes. Economics acknowledges that production of useful goods and services can create problems of environmental pollution. It explores the question of how investing in education helps to develop workers' skills. It probes questions like how to tell when big businesses or big labor unions are operating in a way that benefits society as a whole and when they are operating in a way that benefits their owners or members at the expense of others. It looks at how government spending, taxes, and regulations affect decisions about production and consumption.

It should be clear by now that economics covers a lot of ground. That ground can be divided into two parts: **Microeconomics** focuses on the actions of individual agents within the economy, like households, workers, and businesses; **Macroeconomics** looks at the economy as a whole. It focuses on broad issues such as growth of production, the number of unemployed people, the inflationary increase in prices, government deficits, and levels of exports and imports. Microeconomics and macroeconomics are not separate subjects, but rather complementary perspectives on the overall subject of the economy.

To understand why both microeconomic and macroeconomic perspectives are useful, consider the problem of studying a biological ecosystem like a lake. One person who sets out to study the lake might focus on specific topics: certain kinds of algae or plant life; the characteristics of particular fish or snails; or the trees surrounding the lake. Another person might take an overall view and instead consider the entire ecosystem of the lake from top to bottom; what eats what, how the system stays in a rough balance, and what environmental stresses affect this balance. Both approaches are useful, and both examine the same lake, but the viewpoints are different. In a

similar way, both microeconomics and macroeconomics study the same economy, but each has a different viewpoint.

Whether you are looking at lakes or economics, the micro and the macro insights should blend with each other. In studying a lake, the micro insights about particular plants and animals help to understand the overall food chain, while the macro insights about the overall food chain help to explain the environment in which individual plants and animals live.

In economics, the micro decisions of individual businesses are influenced by whether the macroeconomy is healthy; for example, firms will be more likely to hire workers if the overall economy is growing. In turn, the performance of the macroeconomy ultimately depends on the microeconomic decisions made by individual households and businesses.

Microeconomics

What determines how households and individuals spend their budgets? What combination of goods and services will best fit their needs and wants, given the budget they have to spend? How do people decide whether to work, and if so, whether to work full time or part time? How do people decide how much to save for the future, or whether they should borrow to spend beyond their current means?

What determines the products, and how many of each, a firm will produce and sell? What determines what prices a firm will charge? What determines how a firm will produce its products? What determines how many workers it will hire? How will a firm finance its business? When will a firm decide to expand, downsize, or even close? In the microeconomic part of this book, we will learn about the theory of consumer behavior and the theory of the firm.

Macroeconomics

What determines the level of economic activity in a society? In other words, what determines how many goods and services a nation actually produces? What determines how many jobs are available in an economy? What

determines a nation's standard of living? What causes the economy to speed up or slow down? What causes firms to hire more workers or to lay workers off? Finally, what causes the economy to grow over the long term?

An economy's macroeconomic health can be defined by a number of goals: growth in the standard of living, low unemployment, and low inflation, to name the most important. How can macroeconomic policy be used to pursue these goals? **Monetary policy**, which involves policies that affect bank lending, interest rates, and financial capital markets, is conducted by a nation's central bank. For the United States, this is the Federal Reserve. **Fiscal policy**, which involves government spending and taxes, is determined by a nation's legislative body. For the United States, this is the Congress and the executive branch, which originates the federal budget. These are the main tools the government has to work with. Americans tend to expect that government can fix whatever economic problems we encounter, but to what extent is that expectation realistic? These are just some of the issues that will be explored in the macroeconomic chapters of this book.

Key Concepts and Summary

Microeconomics and macroeconomics are two different perspectives on the economy. The microeconomic perspective focuses on parts of the economy: individuals, firms, and industries. The macroeconomic perspective looks at the economy as a whole, focusing on goals like growth in the standard of living, unemployment, and inflation. Macroeconomics has two types of policies for pursuing these goals: monetary policy and fiscal policy.

Self-Check Questions

Exercise:

Problem:

What would be another example of a “system” in the real world that could serve as a metaphor for micro and macroeconomics?

Solution:

There are many physical systems that would work, for example, the study of planets (micro) in the solar system (macro), or solar systems (micro) in the galaxy (macro).

Review Questions**Exercise:****Problem:**

What is the difference between microeconomics and macroeconomics?

Exercise:

Problem: What are examples of individual economic agents?

Exercise:

Problem: What are the three main goals of macroeconomics?

Critical Thinking Questions**Exercise:****Problem:**

A balanced federal budget and a balance of trade are considered secondary goals of macroeconomics, while growth in the standard of living (for example) is considered a primary goal. Why do you think that is so?

Exercise:

Problem:

Macroeconomics is an aggregate of what happens at the microeconomic level. Would it be possible for what happens at the macro level to differ from how economic agents would react to some stimulus at the micro level? *Hint:* Think about the behavior of crowds.

Glossary

fiscal policy

economic policies that involve government spending and taxes

macroeconomics

the branch of economics that focuses on broad issues such as growth, unemployment, inflation, and trade balance.

microeconomics

the branch of economics that focuses on actions of particular agents within the economy, like households, workers, and business firms

monetary policy

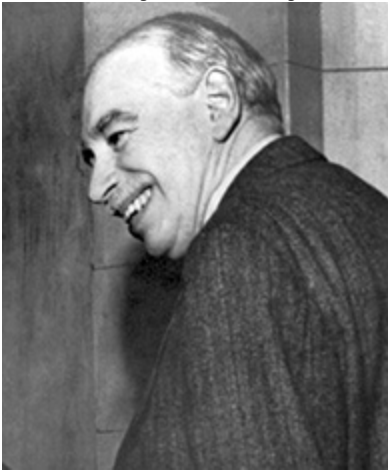
policy that involves altering the level of interest rates, the availability of credit in the economy, and the extent of borrowing

How Economists Use Theories and Models to Understand Economic Issues

By the end of this section, you will be able to:

- Interpret a circular flow diagram
- Explain the importance of economic theories and models
- Describe goods and services markets and labor markets

John Maynard Keynes



One of the most
influential
economists in
modern times was
John Maynard
Keynes. (Credit:
Wikimedia
Commons)

John Maynard Keynes (1883–1946), one of the greatest economists of the twentieth century, pointed out that economics is not just a subject area but also a way of thinking. Keynes, shown in [\[link\]](#), famously wrote in the introduction to a fellow economist's book: “[Economics] is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which

helps its possessor to draw correct conclusions.” In other words, economics teaches you how to think, not what to think.

Note:

Watch this [video](#) about John Maynard Keynes and his influence on economics.

Economists see the world through a different lens than anthropologists, biologists, classicists, or practitioners of any other discipline. They analyze issues and problems with economic theories that are based on particular assumptions about human behavior, that are different than the assumptions an anthropologist or psychologist might use. A **theory** is a simplified representation of how two or more variables interact with each other. The purpose of a theory is to take a complex, real-world issue and simplify it down to its essentials. If done well, this enables the analyst to understand the issue and any problems around it. A good theory is simple enough to be understood, while complex enough to capture the key features of the object or situation being studied.

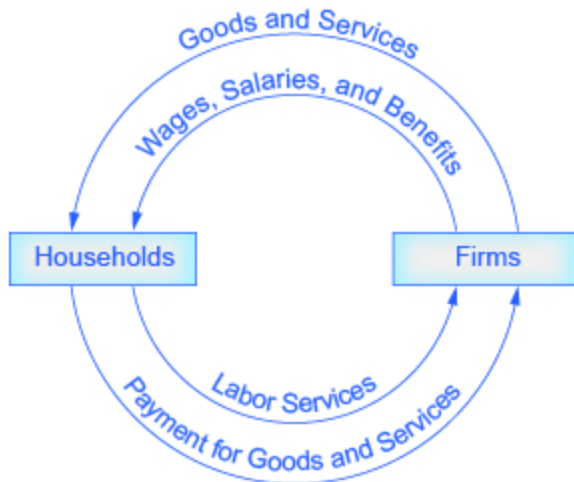
Sometimes economists use the term **model** instead of theory. Strictly speaking, a theory is a more abstract representation, while a model is more applied or empirical representation. Models are used to test theories, but for this course we will use the terms interchangeably.

For example, an architect who is planning a major office building will often build a physical model that sits on a tabletop to show how the entire city block will look after the new building is constructed. Companies often build models of their new products, which are more rough and unfinished than the final product will be, but can still demonstrate how the new product will work.

A good model to start with in economics is the **circular flow diagram**, which is shown in [\[link\]](#). It pictures the economy as consisting of two groups—households and firms—that interact in two markets: the **goods and**

services market in which firms sell and households buy and the **labor market** in which households sell labor to business firms or other employees.

The Circular Flow Diagram



The circular flow diagram shows how households and firms interact in the goods and services market, and in the labor market. The direction of the arrows shows that in the goods and services market, households receive goods and services and pay firms for them. In the labor market, households provide labor and receive payment from firms through wages, salaries, and benefits.

Of course, in the real world, there are many different markets for goods and services and markets for many different types of labor. The circular flow diagram simplifies this to make the picture easier to grasp. In the diagram, firms produce goods and services, which they sell to households in return for revenues. This is shown in the outer circle, and represents the two sides of the product market (for example, the market for goods and services) in which households demand and firms supply. Households sell their labor as workers to firms in return for wages, salaries and benefits. This is shown in

the inner circle and represents the two sides of the labor market in which households supply and firms demand.

This version of the circular flow model is stripped down to the essentials, but it has enough features to explain how the product and labor markets work in the economy. We could easily add details to this basic model if we wanted to introduce more real-world elements, like financial markets, governments, and interactions with the rest of the globe (imports and exports).

Economists carry a set of theories in their heads like a carpenter carries around a toolkit. When they see an economic issue or problem, they go through the theories they know to see if they can find one that fits. Then they use the theory to derive insights about the issue or problem. In economics, theories are expressed as diagrams, graphs, or even as mathematical equations. (Do not worry. In this course, we will mostly use graphs.) Economists do not figure out the answer to the problem first and then draw the graph to illustrate. Rather, they use the graph of the theory to help them figure out the answer. Although at the introductory level, you can sometimes figure out the right answer without applying a model, if you keep studying economics, before too long you will run into issues and problems that you will need to graph to solve. Both micro and macroeconomics are explained in terms of theories and models. The most well-known theories are probably those of supply and demand, but you will learn a number of others.

Key Concepts and Summary

Economists analyze problems differently than do other disciplinary experts. The main tools economists use are economic theories or models. A theory is not an illustration of the answer to a problem. Rather, a theory is a tool for determining the answer.

Self-Check Questions

Exercise:

Problem:

Suppose we extend the circular flow model to add imports and exports. Copy the circular flow diagram onto a sheet of paper and then add a foreign country as a third agent. Draw a rough sketch of the flows of imports, exports, and the payments for each on your diagram.

Solution:

Draw a box outside the original circular flow to represent the foreign country. Draw an arrow from the foreign country to firms, to represent imports. Draw an arrow in the reverse direction representing payments for imports. Draw an arrow from firms to the foreign country to represent exports. Draw an arrow in the reverse direction to represent payments for imports.

Exercise:**Problem:**

What is an example of a problem in the world today, not mentioned in the chapter, that has an economic dimension?

Solution:

There are many such problems. Consider the AIDS epidemic. Why are so few AIDS patients in Africa and Southeast Asia treated with the same drugs that are effective in the United States and Europe? It is because neither those patients nor the countries in which they live have the resources to purchase the same drugs.

Review Questions**Exercise:**

Problem: How did John Maynard Keynes define economics?

Exercise:

Problem:

Are households primarily buyers or sellers in the goods and services market? In the labor market?

Exercise:**Problem:**

Are firms primarily buyers or sellers in the goods and services market? In the labor market?

Critical Thinking Questions**Exercise:****Problem:**

Why is it unfair or meaningless to criticize a theory as “unrealistic?”

Exercise:**Problem:**

Suppose, as an economist, you are asked to analyze an issue unlike anything you have ever done before. Also, suppose you do not have a specific model for analyzing that issue. What should you do? *Hint:* What would a carpenter do in a similar situation?

Glossary

circular flow diagram

a diagram that views the economy as consisting of households and firms interacting in a goods and services market and a labor market

goods and services market

a market in which firms are sellers of what they produce and households are buyers

labor market

the market in which households sell their labor as workers to business firms or other employers

model

see theory

theory

a representation of an object or situation that is simplified while including enough of the key features to help us understand the object or situation

How Economies Can Be Organized: An Overview of Economic Systems

By the end of this section, you will be able to:

- Contrast traditional economies, command economies, and market economies
- Explain gross domestic product (GDP)
- Assess the importance and effects of globalization

Think about what a complex system a modern economy is. It includes all production of goods and services, all buying and selling, all employment. The economic life of every individual is interrelated, at least to a small extent, with the economic lives of thousands or even millions of other individuals. Who organizes and coordinates this system? Who insures that, for example, the number of televisions a society provides is the same as the amount it needs and wants? Who insures that the right number of employees work in the electronics industry? Who insures that televisions are produced in the best way possible? How does it all get done?

There are at least three ways societies have found to organize an economy. The first is the **traditional economy**, which is the oldest economic system and can be found in parts of Asia, Africa, and South America. Traditional economies organize their economic affairs the way they have always done (i.e., tradition). Occupations stay in the family. Most families are farmers who grow the crops they have always grown using traditional methods. What you produce is what you get to consume. Because things are driven by tradition, there is little economic progress or development.

A Command Economy



Ancient Egypt was an example of a command economy. (Credit: Jay Bergesen/Flickr Creative Commons)

Command economies (often called planned economies) are very different. In a **command economy**, economic effort is devoted to goals passed down from a ruler or ruling class. Ancient Egypt was a good example: a large part of economic life was devoted to building pyramids, like those shown in [\[link\]](#), for the pharaohs. Medieval manor life is another example: the lord provided the land for growing crops and protection in the event of war. In return, vassals provided labor and soldiers to do the lord's bidding. In the last century, communism emphasized command economies.

In a command economy, the government decides what goods and services will be produced and what prices will be charged for them. The government decides what methods of production will be used and how much workers will be paid. Many necessities like healthcare and education are provided for free. Currently, Cuba and North Korea have command economies.

A Market Economy



Nothing says
“market” more
than The New
York Stock
Exchange. (Credit:
Erik Drost/Flickr
Creative
Commons)

Although command economies have a very centralized structure for economic decisions, market economies have a very decentralized structure. A **market** is an institution that brings together buyers and sellers of goods or services, who may be either individuals or businesses. The New York Stock Exchange, shown in [\[link\]](#), is a prime example of market in which buyers and sellers are brought together. In a **market economy**, decision-making is decentralized. Market economies are based on **private enterprise**: the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals. Businesses supply goods and services based on demand. (In a command economy, by contrast, resources and businesses are owned by the government.) What goods and services are supplied depends on what is demanded. A person’s income is based on his or her ability to convert resources (especially labor) into something that society values. The more society values the person’s output, the higher the income (think Lady Gaga

or LeBron James). In this scenario, economic decisions are determined by market forces, not governments.

Most economies in the real world are mixed; they combine elements of command and market (and even traditional) systems. The U.S. economy is positioned toward the market-oriented end of the spectrum. Many countries in Europe and Latin America, while primarily market-oriented, have a greater degree of government involvement in economic decisions than does the U.S. economy. China and Russia, while they are closer to having a market-oriented system now than several decades ago, remain closer to the command economy end of the spectrum.

Regulations: The Rules of the Game

Markets and government regulations are always entangled. There is no such thing as an absolutely free market. Regulations always define the “rules of the game” in the economy. Economies that are primarily market-oriented have fewer regulations—ideally just enough to maintain an even playing field for participants. At a minimum, these laws govern matters like safeguarding private property against theft, protecting people from violence, enforcing legal contracts, preventing fraud, and collecting taxes. Conversely, even the most command-oriented economies operate using markets. How else would buying and selling occur? But the decisions of what will be produced and what prices will be charged are heavily regulated. Heavily regulated economies often have **underground economies**, which are markets where the buyers and sellers make transactions without the government’s approval.

The question of how to organize economic institutions is typically not a black-or-white choice between all market or all government, but instead involves a balancing act over the appropriate combination of market freedom and government rules.

Globalization



Cargo ships are one mode of transportation for shipping goods in the global economy.
(Credit: Raul Valdez/Flickr Creative Commons)

The Rise of Globalization

Recent decades have seen a trend toward **globalization**, which is the expanding cultural, political, and economic connections between people around the world. One measure of this is the increased buying and selling of goods, services, and assets across national borders—in other words, international trade and financial capital flows.

Globalization has occurred for a number of reasons. Improvements in shipping, as illustrated by the container ship shown in [\[link\]](#), and air cargo have driven down transportation costs. Innovations in computing and telecommunications have made it easier and cheaper to manage long-distance economic connections of production and sales. Many valuable products and services in the modern economy can take the form of information—for example: computer software; financial advice; travel planning; music, books and movies; and blueprints for designing a building. These products and many others can be transported over telephones and

computer networks at ever-lower costs. Finally, international agreements and treaties between countries have encouraged greater trade.

[\[link\]](#) presents one measure of globalization. It shows the percentage of domestic economic production that was exported for a selection of countries from 2010 to 2013, according to an entity known as The World Bank. **Exports** are the goods and services that are produced domestically and sold abroad. **Imports** are the goods and services that are produced abroad and then sold domestically. The size of total production in an economy is measured by the **gross domestic product (GDP)**. Thus, the ratio of exports divided by GDP measures what share of a country's total economic production is sold in other countries.

Country	2010	2011	2012	2013
Higher Income Countries				
United States	12.4	13.6	13.6	13.5
Belgium	76.2	81.4	82.2	82.8
Canada	29.1	30.7	30.0	30.1
France	26.0	27.8	28.1	28.3
Middle Income Countries				
Brazil	10.9	11.9	12.6	12.6
Mexico	29.9	31.2	32.6	31.7
South Korea	49.4	55.7	56.3	53.9

Country	2010	2011	2012	2013
Lower Income Countries				
Chad	36.8	38.9	36.9	32.2
China	29.4	28.5	27.3	26.4
India	22.0	23.9	24.0	24.8
Nigeria	25.3	31.3	31.4	18.0

The Extent of Globalization (exports/GDP)(Source: <http://databank.worldbank.org/data/>)

In recent decades, the export/GDP ratio has generally risen, both worldwide and for the U.S. economy. Interestingly, the share of U.S. exports in proportion to the U.S. economy is well below the global average, in part because large economies like the United States can contain more of the division of labor inside their national borders. However, smaller economies like Belgium, Korea, and Canada need to trade across their borders with other countries to take full advantage of division of labor, specialization, and economies of scale. In this sense, the enormous U.S. economy is less affected by globalization than most other countries.

[\[link\]](#) also shows that many medium and low income countries around the world, like Mexico and China, have also experienced a surge of globalization in recent decades. If an astronaut in orbit could put on special glasses that make all economic transactions visible as brightly colored lines and look down at Earth, the astronaut would see the planet covered with connections.

So, hopefully, you now have an idea of what economics is about. Before you move to any other chapter of study, be sure to read the very important appendix to this chapter called [The Use of Mathematics in Principles of Economics](#). It is essential that you learn more about how to read and use models in economics.

Note:**Decisions ... Decisions in the Social Media Age**

The world we live in today provides nearly instant access to a wealth of information. Consider that as recently as the late 1970s, the Farmer's Almanac, along with the Weather Bureau of the U.S. Department of Agriculture, were the primary sources American farmers used to determine when to plant and harvest their crops. Today, farmers are more likely to access, online, weather forecasts from the National Oceanic and Atmospheric Administration or watch the Weather Channel. After all, knowing the upcoming forecast could drive when to harvest crops. Consequently, knowing the upcoming weather could change the amount of crop harvested.

Some relatively new information forums, such as Facebook, are rapidly changing how information is distributed; hence, influencing decision making. In 2014, the Pew Research Center reported that 71% of online adults use Facebook. Facebook post topics range from the National Basketball Association, to celebrity singers and performers, to farmers. Information helps us make decisions. Decisions as simple as what to wear today to how many reporters should be sent to cover a crash. Each of these decisions is an economic decision. After all, resources are scarce. If ten reporters are sent to cover an accident, they are not available to cover other stories or complete other tasks. Information provides the knowledge needed to make the best possible decisions on how to utilize scarce resources. Welcome to the world of economics!

Key Concepts and Summary

Societies can be organized as traditional, command, or market-oriented economies. Most societies are a mix. The last few decades have seen globalization evolve as a result of growth in commercial and financial networks that cross national borders, making businesses and workers from different economies increasingly interdependent.

Self-Check Questions

Exercise:**Problem:**

The chapter defines *private enterprise* as a characteristic of market-oriented economies. What would *public enterprise* be? *Hint:* It is a characteristic of command economies.

Solution:

Public enterprise means the factors of production (resources and businesses) are owned and operated by the government.

Exercise:**Problem:**

Why might Belgium, France, Italy, and Sweden have a higher export to GDP ratio than the United States?

Solution:

The United States is a large country economically speaking, so it has less need to trade internationally than the other countries mentioned. (This is the same reason that France and Italy have lower ratios than Belgium or Sweden.) One additional reason is that each of the other countries is a member of the European Union, where trade between members occurs without barriers to trade, like tariffs and quotas.

Review Questions**Exercise:****Problem:**

What are the three ways that societies can organize themselves economically?

Exercise:

Problem:

What is globalization? How do you think it might have affected the economy over the past decade?

Critical Thinking Questions**Exercise:****Problem:**

Why do you think that most modern countries' economies are a mix of command and market types?

Exercise:**Problem:**

Can you think of ways that globalization has helped you economically? Can you think of ways that it has not?

References

The Heritage Foundation. 2015. "2015 Index of Economic Freedom." Accessed March 11, 2015. <http://www.heritage.org/index/ranking>.

Garling, Caleb. "S.F. plane crash: Reporting, emotions on social media," *The San Francisco Chronicle*. July 7, 2013. <http://www.sfgate.com/news/article/S-F-plane-crash-Reporting-emotions-on-social-4651639.php>.

Irvine, Jessica. "Social Networking Sites are Factories of Modern Ideas." *The Sydney Morning Herald*. November 25, 2011. <http://www.smh.com.au/federal-politics/society-and-culture/social-networking-sites-are-factories-of-modern-ideas-20111124-1nwy3.html#ixzz2YZhPYeME>.

Pew Research Center. 2015. "Social Networking Fact Sheet." Accessed March 11, 2015. <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/>.

The World Bank Group. 2015. "World Data Bank." Accessed March 30, 2014. <http://databank.worldbank.org/data/>.

Glossary

command economy

an economy where economic decisions are passed down from government authority and where resources are owned by the government

exports

products (goods and services) made domestically and sold abroad

globalization

the trend in which buying and selling in markets have increasingly crossed national borders

gross domestic product (GDP)

measure of the size of total production in an economy

imports

products (goods and services) made abroad and then sold domestically

market

interaction between potential buyers and sellers; a combination of demand and supply

market economy

an economy where economic decisions are decentralized, resources are owned by private individuals, and businesses supply goods and services based on demand

private enterprise

system where the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals

traditional economy

typically an agricultural economy where things are done the same as they have always been done

underground economy

a market where the buyers and sellers make transactions in violation of one or more government regulations

Introduction to Choice in a World of Scarcity
class="introduction"
Choices and Tradeoffs

In general, the
higher the degree,
the higher the salary.
So why aren't more
people pursuing
higher degrees? The
short answer:
choices and
tradeoffs. (Credit:
modification of
work by "Jim, the
Photographer"/Flick
r Creative
Commons)



Note:**Choices ... To What Degree?**

In 2015, the median income for workers who hold master's degrees varies from males to females. The average of the two is \$2,951 weekly. Multiply this average by 52 weeks, and you get an average salary of \$153,452.

Compare that to the median weekly earnings for a full-time worker over 25 with no higher than a bachelor's degree: \$1,224 weekly and \$63,648 a year. What about those with no higher than a high school diploma in 2015? They earn just \$664 weekly and \$34,528 over 12 months. In other words, says the Bureau of Labor Statistics (BLS), earning a bachelor's degree boosted salaries 54% over what you would have earned if you had stopped your education after high school. A master's degree yields a salary almost double that of a high school diploma.

Given these statistics, we might expect a lot of people to choose to go to college and at least earn a bachelor's degree. Assuming that people want to improve their material well-being, it seems like they would make those choices that give them the greatest opportunity to consume goods and services. As it turns out, the analysis is not nearly as simple as this. In fact, in 2014, the BLS reported that while almost 88% of the population in the United States had a high school diploma, only 33.6% of 25–65 year olds had bachelor's degrees, and only 7.4% of 25–65 year olds in 2014 had earned a master's.

This brings us to the subject of this chapter: why people make the choices they make and how economists go about explaining those choices.

Note:**Introduction to Choice in a World of Scarcity**

In this chapter, you will learn about:

- How Individuals Make Choices Based on Their Budget Constraint
- The Production Possibilities Frontier and Social Choices
- Confronting Objections to the Economic Approach

You will learn quickly when you examine the relationship between economics and scarcity that choices involve tradeoffs. Every choice has a cost.

In 1968, the Rolling Stones recorded “You Can’t Always Get What You Want.” Economists chuckled, because they had been singing a similar tune for decades. English economist Lionel Robbins (1898–1984), in his *Essay on the Nature and Significance of Economic Science* in 1932, described not always getting what you want in this way:

"The time at our disposal is limited. There are only twenty-four hours in the day. We have to choose between the different uses to which they may be put. ... Everywhere we turn, if we choose one thing we must relinquish others which, in different circumstances, we would wish not to have relinquished. Scarcity of means to satisfy given ends is an almost ubiquitous condition of human nature."

Because people live in a world of scarcity, they cannot have all the time, money, possessions, and experiences they wish. Neither can society.

This chapter will continue our discussion of scarcity and the economic way of thinking by first introducing three critical concepts: opportunity cost, marginal decision making, and diminishing returns. Later, it will consider whether the economic way of thinking accurately describes either how choices *are* made or how they *should* be made.

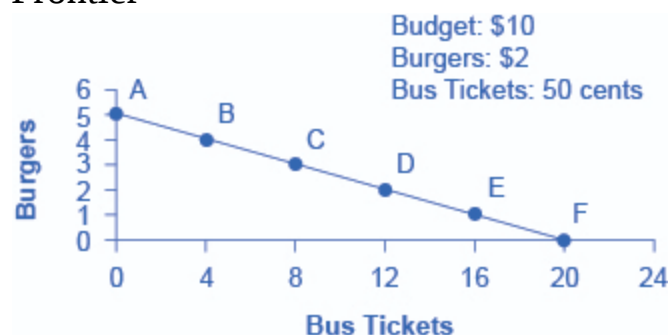
How Individuals Make Choices Based on Their Budget Constraint

By the end of this section, you will be able to:

- Calculate and graph budgets constraints
- Explain opportunity sets and opportunity costs
- Evaluate the law of diminishing marginal utility
- Explain how marginal analysis and utility influence choices

Consider the typical consumer's budget problem. Consumers have a limited amount of income to spend on the things they need and want. Suppose Alphonso has \$10 in spending money each week that he can allocate between bus tickets for getting to work and the burgers that he eats for lunch. Burgers cost \$2 each, and bus tickets are 50 cents each. [\[link\]](#) shows Alphonso's **budget constraint**, that is, the outer boundary of his **opportunity set**. The opportunity set identifies all the opportunities for spending within his budget. The budget constraint indicates all the combinations of burgers and bus tickets Alphonso can afford when he exhausts his budget, given the prices of the two goods. (There are actually many different kinds of budget constraints. You will learn more about them in the chapter on [Consumer Choices](#).)

The Budget Constraint: Alphonso's Consumption Choice Opportunity Frontier



Each point on the budget constraint represents a combination of burgers and bus tickets whose total cost adds up to Alphonso's budget of \$10. The slope of the budget constraint is determined by the

relative price of burgers and bus tickets. All along the budget set, giving up one burger means gaining four bus tickets.

The vertical axis in the figure shows burger purchases and the horizontal axis shows bus ticket purchases. If Alphonso spends all his money on burgers, he can afford five per week. ($\$10 \text{ per week} / \$2 \text{ per burger} = 5 \text{ burgers per week}$.) But if he does this, he will not be able to afford any bus tickets. This choice (zero bus tickets and five burgers) is shown by point A in the figure. Alternatively, if Alphonso spends all his money on bus tickets, he can afford 20 per week. ($\$10 \text{ per week} / \$0.50 \text{ per bus ticket} = 20 \text{ bus tickets per week}$.) Then, however, he will not be able to afford any burgers. This alternative choice (20 bus tickets and zero burgers) is shown by point F.

If Alphonso is like most people, he will choose some combination that includes both bus tickets and burgers. That is, he will choose some combination on the budget constraint that connects points A and F. Every point on (or inside) the constraint shows a combination of burgers and bus tickets that Alphonso can afford. Any point outside the constraint is not affordable, because it would cost more money than Alphonso has in his budget.

The budget constraint clearly shows the tradeoff Alphonso faces in choosing between burgers and bus tickets. Suppose he is currently at point D, where he can afford 12 bus tickets and two burgers. What would it cost Alphonso for one more burger? It would be natural to answer \$2, but that's not the way economists think. Instead they ask, how many bus tickets would Alphonso have to give up to get one more burger, while staying within his budget? The answer is four bus tickets. That is the true cost to Alphonso of one more burger.

The Concept of Tradeoffs and Opportunity Costs

Tradeoffs are choices that face us: to choose to do one thing, means that we give up doing the other things. As an example, you have 3 hours of free time on Thursday. There are several things you could do with that time: watch golf on TV, message your friends on social media, go to a restaurant and have a nice meal, or study economics. Choosing to do one of these involves the tradeoff that you cannot do the other things.

Economists use the term **opportunity cost** to indicate the benefits that must be given up to obtain something that is desired. The idea behind opportunity cost is that the cost of one item is the lost opportunity to do or consume something else; in short, opportunity cost is the value (set of benefits) of the next best alternative. For Alphonso, the opportunity cost of a burger is the four bus tickets he would have to give up. He would decide whether or not to choose the burger depending on whether the value of the burger exceeds the value of the forgone alternative—in this case, bus tickets. A question Alphonso will face is whether the benefit of eating is greater than the benefit of using the bus tickets: choosing one sacrifices the benefits of the other choice. Since people must choose, they inevitably face tradeoffs in which they have to give up things they desire to get other things they desire more.

A fundamental principle of economics is that every choice has an opportunity cost. If you sleep through your economics class (not recommended, by the way), the opportunity cost is the learning you miss from not attending class. If you spend your income on video games, you cannot spend it on movies. If you choose to marry one person, you give up the opportunity to marry anyone else. In short, opportunity cost is all around us and part of human existence.

Identifying Opportunity Cost

In many cases, it is tempting to refer to the opportunity cost as the price. If your cousin buys a new bicycle for \$300, then \$300 measures the amount of “other consumption” that he has given up. However, the opportunity cost represents not only a dollar amount, but also the benefits that would have been realized with the choice not taken. The \$300 could have been used to buy a new computer, allowing your cousin to take online courses at home (a

benefit). Having a bicycle is great for having fun and getting exercise (benefits). If your cousin buys the bicycle, the opportunity costs include not being able to take online courses from home, and if your cousin buys the computer, the opportunity costs include a fun way to get exercise. This problem can loom especially large when costs of time are involved.

For example, consider a boss who decides that all employees will attend a two-day retreat to “build team spirit.” The out-of-pocket monetary cost of the event may involve hiring an outside consulting firm to run the retreat, as well as room and board for all participants. But an opportunity cost exists as well: during the two days of the retreat, none of the employees are doing any other work.

Attending college is another case where the opportunity cost exceeds the monetary cost. The out-of-pocket costs of attending college include tuition, books, room and board, and other expenses. But in addition, during the hours that you are attending class and studying, it is impossible to work at a paying job. Thus, college imposes both an out-of-pocket cost and an opportunity cost of lost earnings.

For clarity, when we refer to tradeoffs or opportunity costs, we want to identify the owner of the decision or choice, and the alternatives involved. As an example, the tradeoff that Sara faced on Thursday afternoon was to either study economics, watch golf on TV, or go to her favorite restaurant for a nice meal. By choosing to watch golf, she gave up the benefits of the next best alternative (studying economics). Her opportunity costs in choosing golf over economics including not doing well on the next test, and ending the semester with a lower grade than she'd hoped.

In some cases, realizing the opportunity cost can alter behavior. Imagine, for example, that you spend \$8 on lunch every day at work. You may know perfectly well that bringing a lunch from home would cost only \$3 a day, so the opportunity cost of buying lunch at the restaurant is \$5 each day (that is, the \$8 buying lunch costs minus the \$3 your lunch from home would cost). \$5 each day does not seem to be that much. However, if you project what that adds up to in a year— $250 \text{ days a year} \times \$5 \text{ per day} = \$1,250$, the cost, perhaps, of a decent vacation. If the opportunity cost is described as “a nice vacation” instead of “\$5 a day,” you might make different choices.

When we consider opportunity costs, we want to remember to include the benefits that were given up in the choice, not just the dollars involved.

Marginal Decision-Making and Diminishing Marginal Utility

The budget constraint framework helps to emphasize that most choices in the real world are not about getting all of one thing or all of another; that is, they are not about choosing either the point at one end of the budget constraint or else the point all the way at the other end. Instead, most choices involve **marginal analysis**, which means comparing the benefits and costs of choosing a little more or a little less of a good.

People desire goods and services for the satisfaction or **utility** those goods and services provide. Utility, as we will see in the chapter on [Consumer Choices](#), is subjective but that does not make it less real. Economists typically assume that the more of some good one consumes (for example, slices of pizza), the more utility one obtains. At the same time, the utility a person receives from consuming the first unit of a good is typically more than the utility received from consuming the fifth or the tenth unit of that same good. When Alphonso chooses between burgers and bus tickets, for example, the first few bus rides that he chooses might provide him with a great deal of utility—perhaps they help him get to a job interview or a doctor’s appointment. But later bus rides might provide much less utility—they may only serve to kill time on a rainy day. Similarly, the first burger that Alphonso chooses to buy may be on a day when he missed breakfast and is ravenously hungry. However, if Alphonso has a burger every single day, the last few burgers may taste pretty boring. The general pattern that consumption of the first few units of any good tends to bring a higher level of utility to a person than consumption of later units is a common pattern. Economists refer to this pattern as the **law of diminishing marginal utility**, which means that as a person receives more of a good, the additional (or marginal) utility from each additional unit of the good declines. In other words, the first slice of pizza brings more satisfaction than the sixth.

The law of diminishing marginal utility explains why people and societies rarely make all-or-nothing choices. You would not say, “My favorite food is ice cream, so I will eat nothing but ice cream from now on.” Instead, even if

you get a very high level of utility from your favorite food, if you ate it exclusively, the additional or marginal utility from those last few servings would not be very high. Similarly, most workers do not say: “I enjoy leisure, so I’ll never work.” Instead, workers recognize that even though some leisure is very nice, a combination of all leisure and no income is not so attractive. The budget constraint framework suggests that when people make choices in a world of scarcity, they will use marginal analysis and think about whether they would prefer a little more or a little less.

Sunk Costs

In the budget constraint framework, all decisions involve what will happen next: that is, what quantities of goods will you consume, how many hours will you work, or how much will you save. These decisions do not look back to past choices. Thus, the budget constraint framework assumes that **sunk costs**, which are costs that were incurred in the past and cannot be recovered, should not affect the current decision.

Consider the case of Selena, who pays \$8 to see a movie, but after watching the film for 30 minutes, she knows that it is truly terrible. Should she stay and watch the rest of the movie because she paid for the ticket, or should she leave? The money she spent is a sunk cost, and unless the theater manager is feeling kindly, Selena will not get a refund. But staying in the movie still means paying an opportunity cost in time. Her choice is whether to spend the next 90 minutes suffering through a cinematic disaster or to do something—anything—else. The lesson of sunk costs is to forget about the money and time that is irretrievably gone and instead to focus on the marginal costs and benefits of current and future options.

For people and firms alike, dealing with sunk costs can be frustrating. It often means admitting an earlier error in judgment. Many firms, for example, find it hard to give up on a new product that is doing poorly because they spent so much money in creating and launching the product. But the lesson of sunk costs is to ignore them and make decisions based on what will happen in the future.

From a Model with Two Goods to One of Many Goods

The budget constraint diagram containing just two goods, like most models used in this book, is not realistic. After all, in a modern economy people choose from thousands of goods. However, thinking about a model with many goods is a straightforward extension of what we discussed here. Instead of drawing just one budget constraint, showing the tradeoff between two goods, you can draw multiple budget constraints, showing the possible tradeoffs between many different pairs of goods. Or in more advanced classes in economics, you would use mathematical equations that include many possible goods and services that can be purchased, together with their quantities and prices, and show how the total spending on all goods and services is limited to the overall budget available. The graph with two goods that was presented here clearly illustrates that every choice has an opportunity cost, which is the point that does carry over to the real world.

Key Concepts and Summary

Economists see the real world as one of scarcity: that is, a world in which people's desires exceed what is possible. As a result, economic behavior involves tradeoffs in which individuals, firms, and society must give up something that they desire to obtain things that they desire more. Individuals face the tradeoff of what quantities of goods and services to consume. The budget constraint, which is the frontier of the opportunity set, illustrates the range of choices available. The slope of the budget constraint is determined by the relative price of the choices. Choices beyond the budget constraint are not affordable.

Opportunity cost measures cost by what is given up in exchange. Sometimes opportunity cost can be measured in money, but it is often useful to consider time as well, or to measure it in terms of the actual resources that must be given up.

Most economic decisions and tradeoffs are not all-or-nothing. Instead, they involve marginal analysis, which means they are about decisions on the margin, involving a little more or a little less. The law of diminishing marginal utility points out that as a person receives more of something—whether it is a specific good or another resource—the additional marginal gains tend to become smaller. Because sunk costs occurred in the past and

cannot be recovered, they should be disregarded in making current decisions.

Self-Check Questions

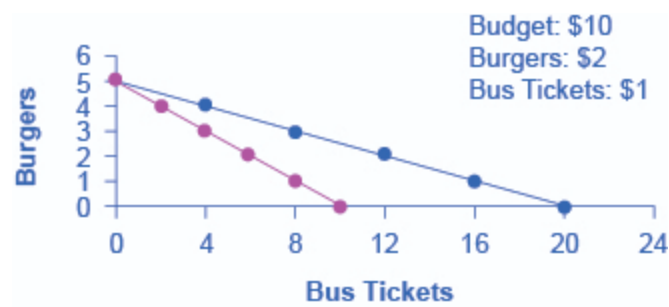
Exercise:

Problem:

Suppose Alphonso's town raised the price of bus tickets to \$1 per trip (while the price of burgers stayed at \$2 and his budget remained \$10 per week.) Draw Alphonso's new budget constraint. What happens to the opportunity cost of bus tickets?

Solution:

The opportunity cost of bus tickets is the number of burgers that must be given up to obtain one more bus ticket. Originally, when the price of bus tickets was 50 cents per trip, this opportunity cost was $0.50/2 = .25$ burgers. The reason for this is that at the original prices, one burger (\$2) costs the same as four bus tickets (\$0.50), so the opportunity cost of a burger is four bus tickets, and the opportunity cost of a bus ticket is .25 (the inverse of the opportunity cost of a burger). With the new, higher price of bus tickets, the opportunity cost rises to $\$1/\2 or 0.50. You can see this graphically since the slope of the new budget constraint is flatter than the original one. If Alphonso spends all of his budget on burgers, the higher price of bus tickets has no impact so the horizontal intercept of the budget constraint is the same. If he spends all of his budget on bus tickets, he can now afford only half as many, so the vertical intercept is half as much. In short, the budget constraint rotates clockwise around the horizontal intercept, flattening as it goes and the opportunity cost of bus tickets increases.



Review Questions

Exercise:

Problem: Explain why scarcity leads to tradeoffs.

Exercise:

Problem:

Explain why individuals make choices that are directly on the budget constraint, rather than inside the budget constraint or outside it.

Critical Thinking Question

Exercise:

Problem:

Suppose Alphonso's town raises the price of bus tickets from \$0.50 to \$1 and the price of burgers rises from \$2 to \$4. Why is the opportunity cost of bus tickets unchanged? Suppose Alphonso's weekly spending money increases from \$10 to \$20. How is his budget constraint affected from all three changes? Explain.

Problems

Use this information to answer the following 4 questions: Marie has a weekly budget of \$24, which she likes to spend on magazines and pies.

Exercise:**Problem:**

If the price of a magazine is \$4 each, what is the maximum number of magazines she could buy in a week?

Exercise:**Problem:**

If the price of a pie is \$12, what is the maximum number of pies she could buy in a week?

Exercise:**Problem:**

Draw Marie's budget constraint with pies on the horizontal axis and magazines on the vertical axis. What is the slope of the budget constraint?

Exercise:

Problem: What is Marie's opportunity cost of purchasing a pie?

References

Bureau of Labor Statistics, U.S. Department of Labor. 2015. "Median Weekly Earnings by Educational Attainment in 2014." Accessed March 27, 2015. <http://www.bls.gov/opub/ted/2015/median-weekly-earnings-by-education-gender-race-and-ethnicity-in-2014.htm>.

Robbins, Lionel. *An Essay on the Nature and Significance of Economic Science*. London: Macmillan. 1932.

United States Department of Transportation. "Total Passengers on U.S Airlines and Foreign Airlines U.S. Flights Increased 1.3% in 2012 from 2011." Accessed October 2013. http://www.rita.dot.gov/bts/press_releases/bts016_13

Glossary

budget constraint

all possible consumption combinations of goods that someone can afford, given the prices of goods, when all income is spent; the boundary of the opportunity set

law of diminishing marginal utility

as we consume more of a good or service, the utility we get from additional units of the good or service tend to become smaller than what we received from earlier units

marginal analysis

examination of decisions on the margin, meaning a little more or a little less from the status quo

opportunity cost

measures cost by what is given up in exchange; opportunity cost measures the value of the forgone alternative

opportunity set

all possible combinations of consumption that someone can afford given the prices of goods and the individual's income

sunk costs

costs that are made in the past and cannot be recovered

utility

satisfaction, usefulness, or value one obtains from consuming goods and services

The Production Possibilities Frontier and Social Choices

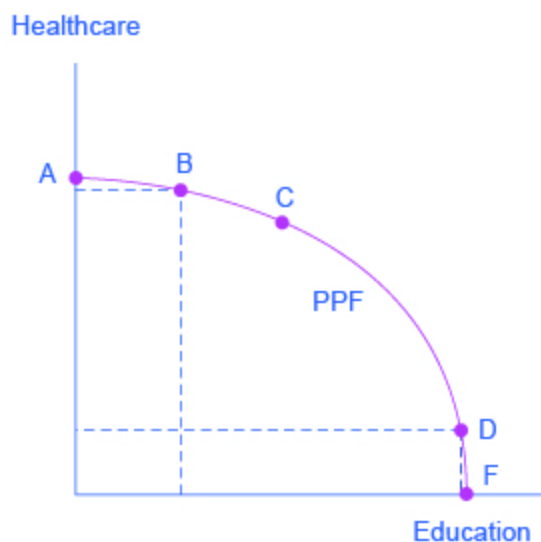
By the end of this section, you will be able to:

- Interpret production possibilities frontier graphs
- Contrast a budget constraint and a production possibilities frontier
- Explain the relationship between a production possibilities frontier and the law of diminishing returns
- Contrast productive efficiency and allocative efficiency
- Define comparative advantage

Just as individuals cannot have everything they want and must instead make choices, society as a whole cannot have everything it might want, either. This section of the chapter will explain the constraints faced by society, using a model called the **production possibilities frontier (PPF)**. There are more similarities than differences between individual choice and social choice. As you read this section, focus on the similarities.

Because society has limited resources (e.g., labor, land, capital, raw materials) at any point in time, there is a limit to the quantities of goods and services it can produce. Suppose a society desires two products, healthcare and education. This situation is illustrated by the production possibilities frontier in [\[link\]](#).

A Healthcare vs. Education Production Possibilities Frontier



This production possibilities frontier shows a tradeoff between devoting social resources to healthcare and devoting them to education. At A all resources go to healthcare and at B, most go to healthcare. At D most resources go to education, and at F, all go to education.

In [\[link\]](#), healthcare is shown on the vertical axis and education is shown on the horizontal axis. If the society were to allocate all of its resources to healthcare, it could produce at point A. But it would not have any resources to produce education. If it were to allocate all of its resources to education, it could produce at point F. Alternatively, the society could choose to produce any combination of healthcare and education shown on the production possibilities frontier. In effect, the production possibilities frontier plays the same role for society as the budget constraint plays for Alphonso. Society can choose any combination of the two goods on or inside the PPF. But it does not have enough resources to produce outside the PPF.

Most important, the production possibilities frontier clearly shows the tradeoff between healthcare and education. Suppose society has chosen to operate at point B, and it is considering producing more education. Because the PPF is downward sloping from left to right, the only way society can obtain more education is by giving up some healthcare. That is the tradeoff society faces. Suppose it considers moving from point B to point C. What would the opportunity cost be for the additional education? The opportunity cost would be the healthcare society has to give up. Just as with Alphonso's budget constraint, the opportunity cost is shown by the slope of the production possibilities frontier. By now you might be saying, "Hey, this PPF is sounding like the budget constraint." If so, read the following Clear It Up feature.

Note:

What's the difference between a budget constraint and a PPF?

There are two major differences between a budget constraint and a production possibilities frontier. The first is the fact that the budget constraint is a straight line. This is because its slope is given by the relative prices of the two goods. In contrast, the PPF has a curved shape because of the law of the diminishing returns. The second is the absence of specific numbers on the axes of the PPF. There are no specific numbers because we do not know the exact amount of resources this imaginary economy has, nor do we know how many resources it takes to produce healthcare and how many resources it takes to produce education. If this were a real world example, that data would be available. An additional reason for the lack of numbers is that there is no single way to measure levels of education and healthcare. However, when you think of improvements in education, you can think of accomplishments like more years of school completed, fewer high-school dropouts, and higher scores on standardized tests. When you think of improvements in healthcare, you can think of longer life expectancies, lower levels of infant mortality, and fewer outbreaks of disease.

Whether or not we have specific numbers, conceptually we can measure the opportunity cost of additional education as society moves from point B to point C on the PPF. The additional education is measured by the horizontal distance between B and C. The foregone healthcare is given by the vertical distance between B and C. The slope of the PPF between B and C is (approximately) the vertical distance (the “rise”) over the horizontal distance (the “run”). This is the opportunity cost of the additional education.

The Shape of the PPF and the Law of Diminishing Returns

The budget constraints presented earlier in this chapter, showing individual choices about what quantities of goods to consume, were all straight lines. The reason for these straight lines was that the slope of the budget constraint was determined by relative prices of the two goods in the consumption budget constraint. However, the production possibilities

frontier for healthcare and education was drawn as a curved line. Why does the PPF have a different shape?

To understand why the PPF is curved, start by considering point A at the top left-hand side of the PPF. At point A, all available resources are devoted to healthcare and none are left for education. This situation would be extreme and even ridiculous. For example, children are seeing a doctor every day, whether they are sick or not, but not attending school. People are having cosmetic surgery on every part of their bodies, but no high school or college education exists. Now imagine that some of these resources are diverted from healthcare to education, so that the economy is at point B instead of point A. Diverting some resources away from A to B causes relatively little reduction in health because the last few marginal dollars going into healthcare services are not producing much additional gain in health. However, putting those marginal dollars into education, which is completely without resources at point A, can produce relatively large gains. For this reason, the shape of the PPF from A to B is relatively flat, representing a relatively small drop-off in health and a relatively large gain in education.

Now consider the other end, at the lower right, of the production possibilities frontier. Imagine that society starts at choice D, which is devoting nearly all resources to education and very few to healthcare, and moves to point F, which is devoting *all* spending to education and none to healthcare. For the sake of concreteness, you can imagine that in the movement from D to F, the last few doctors must become high school science teachers, the last few nurses must become school librarians rather than dispensers of vaccinations, and the last few emergency rooms are turned into kindergartens. The gains to education from adding these last few resources to education are very small. However, the opportunity cost lost to health will be fairly large, and thus the slope of the PPF between D and F is steep, showing a large drop in health for only a small gain in education.

The lesson is not that society is likely to make an extreme choice like devoting no resources to education at point A or no resources to health at point F. Instead, the lesson is that the gains from committing additional marginal resources to education depend on how much is already being

spent. If on the one hand, very few resources are currently committed to education, then an increase in resources used can bring relatively large gains. On the other hand, if a large number of resources are already committed to education, then committing additional resources will bring relatively smaller gains.

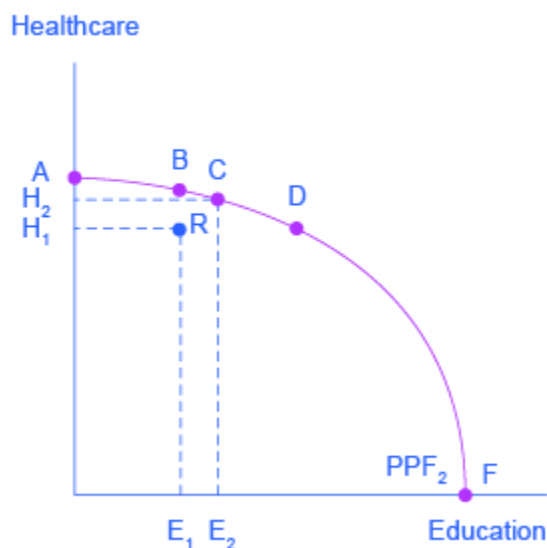
This pattern is common enough that it has been given a name: the **law of diminishing returns**, which holds that as additional increments of resources are added to a certain purpose, the marginal benefit from those additional increments will decline. When government spends a certain amount more on reducing crime, for example, the original gains in reducing crime could be relatively large. But additional increases typically cause relatively smaller reductions in crime, and paying for enough police and security to reduce crime to nothing at all would be tremendously expensive.

The curvature of the production possibilities frontier shows that as additional resources are added to education, moving from left to right along the horizontal axis, the original gains are fairly large, but gradually diminish. Similarly, as additional resources are added to healthcare, moving from bottom to top on the vertical axis, the original gains are fairly large, but again gradually diminish. In this way, the law of diminishing returns produces the outward-bending shape of the production possibilities frontier.

Productive Efficiency and Allocative Efficiency

The study of economics does not presume to tell a society what choice it should make along its production possibilities frontier. In a market-oriented economy with a democratic government, the choice will involve a mixture of decisions by individuals, firms, and government. However, economics can point out that some choices are unambiguously better than others. This observation is based on the concept of efficiency. In everyday usage, efficiency refers to lack of waste. An inefficient machine operates at high cost, while an efficient machine operates at lower cost, because it is not wasting energy or materials. An inefficient organization operates with long delays and high costs, while an efficient organization meets schedules, is focused, and performs within budget.

The production possibilities frontier can illustrate two kinds of efficiency: productive efficiency and allocative efficiency. [\[link\]](#) illustrates these ideas using a production possibilities frontier between healthcare and education. Productive and Allocative Efficiency



Productive efficiency means it is impossible to produce more of one good without decreasing the quantity that is produced of another good. Thus, all choices along a given PPF like B, C, and D display productive efficiency, but R does not. Allocative efficiency means that the particular mix of goods being produced—that is, the specific choice along the production possibilities frontier—represents the allocation that society most desires.

Productive efficiency means that, given the available inputs and technology, it is impossible to produce more of one good without decreasing the quantity that is produced of another good. All choices on the

PPF in [\[link\]](#), including A, B, C, D, and F, display productive efficiency. As a firm moves from any one of these choices to any other, either healthcare increases and education decreases or vice versa. However, any choice inside the production possibilities frontier is productively inefficient and wasteful because it is possible to produce more of one good, the other good, or some combination of both goods.

For example, point R is productively inefficient because it is possible at choice C to have more of both goods: education on the horizontal axis is higher at point C than point R (E_2 is greater than E_1), and healthcare on the vertical axis is also higher at point C than point R (H_2 is greater than H_1).

The particular mix of goods and services being produced—that is, the specific combination of healthcare and education chosen along the production possibilities frontier—can be shown as a ray (line) from the origin to a specific point on the PPF. Output mixes that had more healthcare (and less education) would have a steeper ray, while those with more education (and less healthcare) would have a flatter ray.

Allocative efficiency means that the particular mix of goods a society produces represents the combination that society most desires. How to determine what a society desires can be a controversial question, and is usually discussed in political science, sociology, and philosophy classes as well as in economics. At its most basic, allocative efficiency means producers supply the quantity of each product that consumers demand. Only one of the productively efficient choices will be the allocatively efficient choice for society as a whole.

Why Society Must Choose

Every economy faces two situations in which it may be able to expand consumption of all goods. In the first case, a society may discover that it has been using its resources inefficiently, in which case by improving efficiency and producing on the production possibilities frontier, it can have more of all goods (or at least more of some and less of none). In the second case, as resources grow over a period of years (e.g., more labor and more capital), the economy grows. As it does, the production possibilities frontier

for a society will tend to shift outward and society will be able to afford more of all goods.

So what else may contribute to the growth in the size of an economy? Put another way, since the PPF (Production Possibilities Frontier) illustrates the capacity of an economy to produce, what changes enable the capacity to get bigger? Remember, the factors of production are land, labor and capital, and the PPF describes the capacity of the economy using these factors in an efficient way.

To expand the capacity of an economy requires a change in either the quantity or the quality of at least one factor of production. What does this mean? For labor, it means that increasing the quantity requires either an increase in births or an increase in immigration. Increasing the quality of labor? Well, you are being educated, and that is an improvement in your capacity to work. The same holds true for other types of training, such as classes held by the employer, as well as degrees earned from a college or university.

Capital can be improved as well. Improving the quality of capital often is accomplished by the creation of a technological change or advancement (e.g., introducing artificial intelligence into robots at a factory). The quantity of capital can be increased, by attracting foreign investments in plant and equipment (think of a foreign based car brand, opening a factory in the U.S.).

Land can be improved as well; it's easier to improve the quality (e.g., irrigation of crop land), than it is to improved the quantity (not too easy to conquer part of another country, right?). That said, the two areas of improvement that generally get the most attention in the U.S. seem to be the improvement in the quality of labor and of capital (education and technology). Evidence is pretty available in most political election cycles: most candidates for any elected office, provide their views on education and technology, and almost always advocate for advances in both. There's a benefit to most all participants households (workers), firms (your employer) and the economy overall.

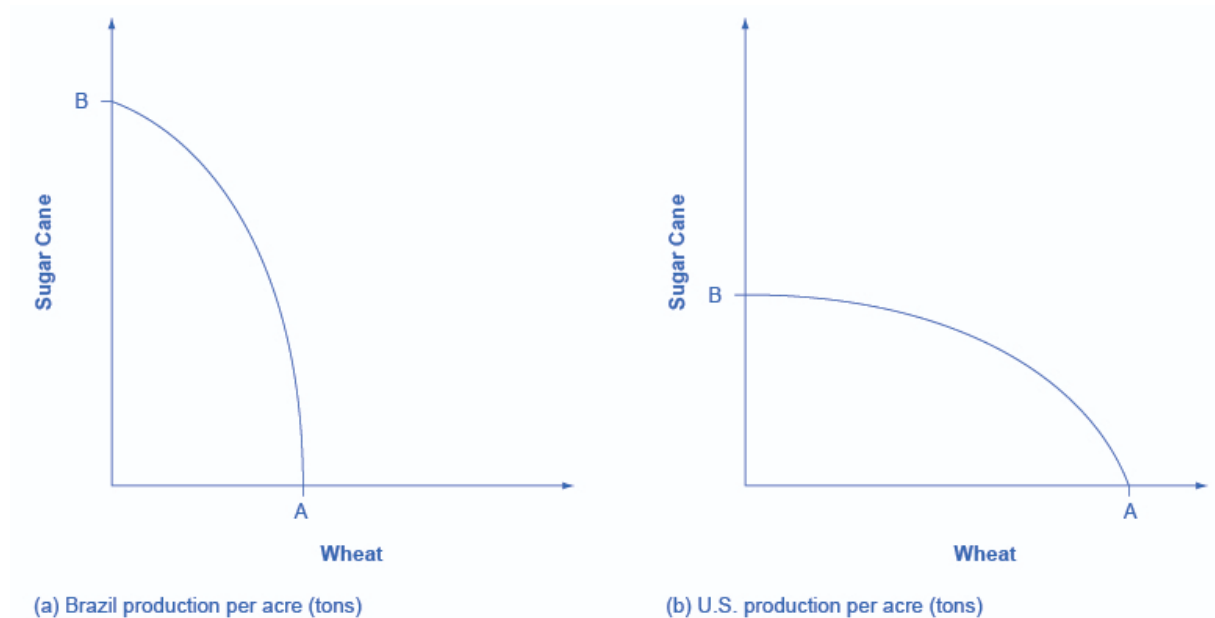
But improvements in productive efficiency take time to discover and implement, and economic growth happens only gradually. So, a society must choose between tradeoffs in the present. For government, this process often involves trying to identify where additional spending could do the most good and where reductions in spending would do the least harm. At the individual and firm level, the market economy coordinates a process in which firms seek to produce goods and services in the quantity, quality, and price that people want. But for both the government and the market economy in the short term, increases in production of one good typically mean offsetting decreases somewhere else in the economy.

The PPF and Comparative Advantage

While every society must choose how much of each good it should produce, it does not need to produce every single good it consumes. Often how much of a good a country decides to produce depends on how expensive it is to produce it versus buying it from a different country. As we saw earlier, the curvature of a country's PPF gives us information about the tradeoff between devoting resources to producing one good versus another. In particular, its slope gives the opportunity cost of producing one more unit of the good in the x-axis in terms of the other good (in the y-axis). Countries tend to have different opportunity costs of producing a specific good, either because of different climates, geography, technology or skills.

Suppose two countries, the US and Brazil, need to decide how much they will produce of two crops: sugar cane and wheat. Due to its climatic conditions, Brazil can produce a lot of sugar cane per acre but not much wheat. Conversely, the U.S. can produce a lot of wheat per acre, but not much sugar cane. Clearly, Brazil has a lower opportunity cost of producing sugar cane (in terms of wheat) than the U.S. The reverse is also true; the U.S. has a lower opportunity cost of producing wheat than Brazil. This can be illustrated by the PPFs of the two countries in [\[link\]](#)

Production Possibility Frontier for the U.S. and Brazil



The U.S. PPF is flatter than the Brazil PPF implying that the opportunity cost of wheat in term of sugar cane is lower in the U.S. than in Brazil. Conversely, the opportunity cost of sugar cane is lower in Brazil. The U.S. has comparative advantage in wheat and Brazil has comparative advantage in sugar cane.

When a country can produce a good at a lower opportunity cost than another country, we say that this country has a **comparative advantage** in that good. In our example, Brazil has a comparative advantage in sugar cane and the U.S. has a comparative advantage in wheat. One can easily see this with a simple observation of the extreme production points in the PPFs of the two countries. If Brazil devoted all of its resources to producing wheat, it would be producing at point A. If however it had devoted all of its resources to producing sugar cane instead, it would be producing a much larger amount, at point B. By moving from point A to point B Brazil would give up a relatively small quantity in wheat production to obtain a large production in sugar cane. The opposite is true for the U.S. If the U.S. moved from point A to B and produced only sugar cane, this would result in a large opportunity cost in terms of foregone wheat production.

The slope of the PPF gives the opportunity cost of producing an additional unit of wheat. While the slope is not constant throughout the PPFs, it is quite apparent that the PPF in Brazil is much steeper than in the U.S., and therefore the opportunity cost of wheat generally higher in Brazil. In the chapter on [International Trade](#) you will learn that countries' differences in comparative advantage determine which goods they will choose to produce and trade. When countries engage in trade, they specialize in the production of the goods that they have comparative advantage in, and trade part of that production for goods they do not have comparative advantage in. With trade, goods are produced where the opportunity cost is lowest, so total production increases, benefiting both trading parties.

Key Concepts and Summary

A production possibilities frontier defines the set of choices society faces for the combinations of goods and services it can produce given the resources available. The shape of the PPF is typically curved outward, rather than straight. Choices outside the PPF are unattainable and choices inside the PPF are wasteful. Over time, a growing economy will tend to shift the PPF outwards.

The law of diminishing returns holds that as increments of additional resources are devoted to producing something, the marginal increase in output will become smaller and smaller. All choices along a production possibilities frontier display productive efficiency; that is, it is impossible to use society's resources to produce more of one good without decreasing production of the other good. The specific choice along a production possibilities frontier that reflects the mix of goods society prefers is the choice with allocative efficiency. The curvature of the PPF is likely to differ by country, which results in different countries having comparative advantage in different goods. Total production can increase if countries specialize in the goods they have comparative advantage in and trade some of their production for the remaining goods.

Self-Check Questions

Exercise:

Problem:

Return to the example in [\[link\]](#). Suppose there is an improvement in medical technology that enables more healthcare to be provided with the same amount of resources. How would this affect the production possibilities curve and, in particular, how would it affect the opportunity cost of education?

Solution:

Because of the improvement in technology, the vertical intercept of the PPF would be at a higher level of healthcare. In other words, the PPF would rotate clockwise around the horizontal intercept. This would make the PPF steeper, corresponding to an increase in the opportunity cost of education, since resources devoted to education would now mean forgoing a greater quantity of healthcare.

Exercise:**Problem:**

Could a nation be producing in a way that is allocatively efficient, but productively inefficient?

Solution:

No. Allocative efficiency requires productive efficiency, because it pertains to choices along the production possibilities frontier.

Exercise:**Problem:**

What are the similarities between a consumer's budget constraint and society's production possibilities frontier, not just graphically but analytically?

Solution:

Both the budget constraint and the PPF show the constraint that each operates under. Both show a tradeoff between having more of one good but less of the other. Both show the opportunity cost graphically as the slope of the constraint (budget or PPF).

Review Questions

Exercise:

Problem: What is comparative advantage?

Exercise:

Problem: What does a production possibilities frontier illustrate?

Exercise:

Problem:

Why is a production possibilities frontier typically drawn as a curve, rather than a straight line?

Exercise:

Problem:

Explain why societies cannot make a choice above their production possibilities frontier and should not make a choice below it.

Exercise:

Problem: What are diminishing marginal returns?

Exercise:

Problem: What is productive efficiency? Allocative efficiency?

Critical Thinking Questions

Exercise:**Problem:**

During the Second World War, Germany's factories were decimated. It also suffered many human casualties, both soldiers and civilians. How did the war affect Germany's production possibilities curve?

Exercise:**Problem:**

It is clear that productive inefficiency is a waste since resources are being used in a way that produces less goods and services than a nation is capable of. Why is allocative inefficiency also wasteful?

Glossary

allocative efficiency

when the mix of goods being produced represents the mix that society most desires

comparative advantage

when a country can produce a good at a lower cost in terms of other goods; or, when a country has a lower opportunity cost of production

law of diminishing returns

as additional increments of resources are added to producing a good or service, the marginal benefit from those additional increments will decline

production possibilities frontier (PPF)

a diagram that shows the productively efficient combinations of two products that an economy can produce given the resources it has available.

productive efficiency

when it is impossible to produce more of one good (or service) without decreasing the quantity produced of another good (or service)

Confronting Objections to the Economic Approach

By the end of this section, you will be able to:

- Analyze arguments against economic approaches to decision-making
- Interpret a tradeoff diagram
- Contrast normative statements and positive statements

It is one thing to understand the economic approach to decision-making and another thing to feel comfortable applying it. The sources of discomfort typically fall into two categories: that people do not act in the way that fits the economic way of thinking, and that even if people did act that way, they should try not to. Let's consider these arguments in turn.

First Objection: People, Firms, and Society Do Not Act Like This

The economic approach to decision-making seems to require more information than most individuals possess and more careful decision-making than most individuals actually display. After all, do you or any of your friends draw a budget constraint and mutter to yourself about maximizing utility before you head to the shopping mall? Do members of the U.S. Congress contemplate production possibilities frontiers before they vote on the annual budget? The messy ways in which people and societies operate somehow doesn't look much like neat budget constraints or smoothly curving production possibilities frontiers.

However, the economics approach can be a useful way to analyze and understand the tradeoffs of economic decisions even so. To appreciate this point, imagine for a moment that you are playing basketball, dribbling to the right, and throwing a bounce-pass to the left to a teammate who is running toward the basket. A physicist or engineer could work out the correct speed and trajectory for the pass, given the different movements involved and the weight and bounciness of the ball. But when you are playing basketball, you do not perform any of these calculations. You just pass the ball, and if you are a good player, you will do so with high accuracy.

Someone might argue: “The scientist’s formula of the bounce-pass requires a far greater knowledge of physics and far more specific information about speeds of movement and weights than the basketball player actually has, so it must be an unrealistic description of how basketball passes are actually made.” This reaction would be wrongheaded. The fact that a good player can throw the ball accurately because of practice and skill, without making a physics calculation, does not mean that the physics calculation is wrong.

Similarly, from an economic point of view, someone who goes shopping for groceries every week has a great deal of practice with how to purchase the combination of goods that will provide that person with utility, even if the shopper does not phrase decisions in terms of a budget constraint.

Government institutions may work imperfectly and slowly, but in general, a democratic form of government feels pressure from voters and social institutions to make the choices that are most widely preferred by people in that society. So, when thinking about the economic actions of groups of people, firms, and society, it is reasonable, as a first approximation, to analyze them with the tools of economic analysis. For more on this, read about behavioral economics in the chapter on [Consumer Choices](#).

Second Objection: People, Firms, and Society Should Not Act This Way

The economics approach portrays people as self-interested. For some critics of this approach, even if self-interest is an accurate description of how people behave, these behaviors are not moral. Instead, the critics argue that people should be taught to care more deeply about others. Economists offer several answers to these concerns.

First, economics is not a form of moral instruction. Rather, it seeks to describe economic behavior as it actually exists. Philosophers draw a distinction between **positive statements**, which describe the world as it is, and **normative statements**, which describe how the world should be. For example, an economist could analyze a proposed subway system in a certain city. If the expected benefits exceed the costs, he concludes that the project is worth doing—an example of positive analysis. Another economist argues for extended unemployment compensation during the Great

Depression because a rich country like the United States should take care of its less fortunate citizens—an example of normative analysis.

Even if the line between positive and normative statements is not always crystal clear, economic analysis does try to remain rooted in the study of the actual people who inhabit the actual economy. Fortunately however, the assumption that individuals are purely self-interested is a simplification about human nature. In fact, we need to look no further than to Adam Smith, the very father of modern economics to find evidence of this. The opening sentence of his book, *The Theory of Moral Sentiments*, puts it very clearly: “How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.” Clearly, individuals are both self-interested and altruistic.

Second, self-interested behavior and profit-seeking can be labeled with other names, such as personal choice and freedom. The ability to make personal choices about buying, working, and saving is an important personal freedom. Some people may choose high-pressure, high-paying jobs so that they can earn and spend a lot of money on themselves. Others may earn a lot of money and give it to charity or spend it on their friends and family. Others may devote themselves to a career that can require a great deal of time, energy, and expertise but does not offer high financial rewards, like being an elementary school teacher or a social worker. Still others may choose a job that does not take lots of their time or provide a high level of income, but still leaves time for family, friends, and contemplation. Some people may prefer to work for a large company; others might want to start their own business. People’s freedom to make their own economic choices has a moral value worth respecting.

Note:

Is a diagram by any other name the same?

When you study economics, you may feel buried under an avalanche of diagrams: diagrams in the text, diagrams in the lectures, diagrams in the problems, and diagrams on exams. Your goal should be to recognize the

common underlying logic and pattern of the diagrams, not to memorize each of the individual diagrams.

This chapter uses only one basic diagram, although it is presented with different sets of labels. The consumption budget constraint and the production possibilities frontier for society, as a whole, are the same basic diagram. [\[link\]](#) shows an individual budget constraint and a production possibilities frontier for two goods, Good 1 and Good 2. The tradeoff diagram always illustrates three basic themes: scarcity, tradeoffs, and economic efficiency.

The first theme is scarcity. It is not feasible to have unlimited amounts of both goods. But even if the budget constraint or a PPF shifts, scarcity remains—just at a different level. The second theme is tradeoffs. As depicted in the budget constraint or the production possibilities frontier, it is necessary to give up some of one good to gain more of the other good. The details of this tradeoff vary. In a budget constraint, the tradeoff is determined by the relative prices of the goods: that is, the relative price of two goods in the consumption choice budget constraint. These tradeoffs appear as a straight line. However, the tradeoffs in many production possibilities frontiers are represented by a curved line because the law of diminishing returns holds that as resources are added to an area, the marginal gains tend to diminish. Regardless of the specific shape, tradeoffs remain.

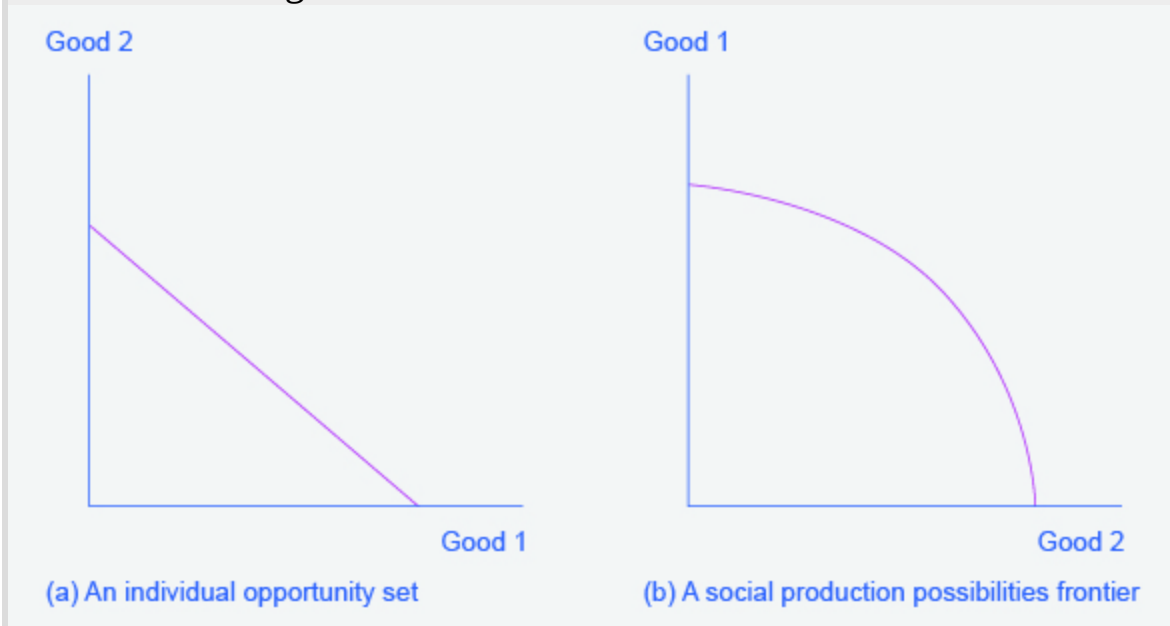
The third theme is economic efficiency, or getting the most benefit from scarce resources. All choices on the production possibilities frontier show productive efficiency because in such cases, there is no way to increase the quantity of one good without decreasing the quantity of the other.

Similarly, when an individual makes a choice along a budget constraint, there is no way to increase the quantity of one good without decreasing the quantity of the other. The choice on a production possibilities set that is socially preferred, or the choice on an individual's budget constraint that is personally preferred, will display allocative efficiency.

The basic budget constraint/production possibilities frontier diagram will recur throughout this book. Some examples include using these tradeoff diagrams to analyze trade, labor supply versus leisure, saving versus consumption, environmental protection and economic output, equality of incomes and economic output, and the macroeconomic tradeoff between consumption and investment. Do not be confused by the different labels.

The budget constraint/production possibilities frontier diagram is always just a tool for thinking carefully about scarcity, tradeoffs, and efficiency in a particular situation.

The Tradeoff Diagram



Both the individual opportunity set (or budget constraint) and the social production possibilities frontier show the constraints under which individual consumers and society as a whole operate. Both diagrams show the tradeoff in choosing more of one good at the cost of less of the other.

Third, self-interested behavior can lead to positive social results. For example, when people work hard to make a living, they create economic output. Consumers who are looking for the best deals will encourage businesses to offer goods and services that meet their needs. Adam Smith, writing in *The Wealth of Nations*, christened this property the **invisible hand**. In describing how consumers and producers interact in a market economy, Smith wrote:

"Every individual...generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain. And he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention...By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it."

The metaphor of the invisible hand suggests the remarkable possibility that broader social good can emerge from selfish individual actions.

Fourth, even people who focus on their own self-interest in the economic part of their life often set aside their own narrow self-interest in other parts of life. For example, you might focus on your own self-interest when asking your employer for a raise or negotiating to buy a car. But then you might turn around and focus on other people when you volunteer to read stories at the local library, help a friend move to a new apartment, or donate money to a charity. Self-interest is a reasonable starting point for analyzing many economic decisions, without needing to imply that people never do anything that is not in their own immediate self-interest.

Note:**Choices ... To What Degree?**

What have we learned? We know that scarcity impacts all the choices we make. So, an economist might argue that people do not go on to get bachelor's degrees or master's degrees because they do not have the resources to make those choices or because their incomes are too low and/or the price of these degrees is too high. A bachelor's degree or a master's degree may not be available in their opportunity set.

The price of these degrees may be too high not only because the actual price, college tuition (and perhaps room and board), is too high. An economist might also say that for many people, the full opportunity cost of a bachelor's degree or a master's degree is too high. For these people, they are unwilling or unable to make the tradeoff of giving up years of working, and earning an income, to earn a degree.

Finally, the statistics introduced at the start of the chapter reveal information about intertemporal choices. An economist might say that people choose not to get a college degree because they may have to borrow money to go to college, and the interest they have to pay on that loan in the future will affect their decisions today. Also, it could be that some people have a preference for current consumption over future consumption, so they choose to work now at a lower salary and consume now, rather than putting that consumption off until after they graduate college.

Key Concepts and Summary

The economic way of thinking provides a useful approach to understanding human behavior. Economists make the careful distinction between positive statements, which describe the world as it is, and normative statements, which describe how the world should be. Even when economics analyzes the gains and losses from various events or policies, and thus draws normative conclusions about how the world should be, the analysis of economics is rooted in a positive analysis of how people, firms, and governments actually behave, not how they should behave.

Self-Check Questions

Exercise:

Problem:

Individuals may not act in the rational, calculating way described by the economic model of decision making, measuring utility and costs at the margin, but can you make a case that they behave approximately that way?

Solution:

When individuals compare cost per unit in the grocery store, or characteristics of one product versus another, they are behaving approximately like the model describes.

Exercise:**Problem:**

Would an op-ed piece in a newspaper urging the adoption of a particular economic policy be considered a positive or normative statement?

Solution:

Since an op-ed makes a case for what should be, it is considered normative.

Exercise:**Problem:**

Would a research study on the effects of soft drink consumption on children's cognitive development be considered a positive or normative statement?

Solution:

Assuming that the study is not taking an explicit position about whether soft drink consumption is good or bad, but just reporting the science, it would be considered positive.

Review Questions**Exercise:****Problem:**

What is the difference between a positive and a normative statement?

Exercise:

Problem:

Is the economic model of decision-making intended as a literal description of how individuals, firms, and the governments actually make decisions?

Exercise:**Problem:**

What are four responses to the claim that people should not behave in the way described in this chapter?

Critical Thinking Questions**Exercise:****Problem:**

What assumptions about the economy must be true for the invisible hand to work? To what extent are those assumptions valid in the real world?

Exercise:**Problem:**

Do economists have any particular expertise at making normative arguments? In other words, they have expertise at making positive statements (i.e., what *will* happen) about some economic policy, for example, but do they have special expertise to judge whether or not the policy *should* be undertaken?

References

Smith, Adam. "Of Restraints upon the Importation from Foreign Countries." In *The Wealth of Nations*. London: Methuen & Co., 1904, first pub 1776), I.V. 2.9.

Smith, Adam. "Of the Propriety of Action." In *The Theory of Moral Sentiments*. London: A. Millar, 1759, 1.

Glossary

invisible hand

idea that self-interested behavior by individuals can lead to positive social outcomes

normative statement

statement which describes how the world should be

positive statement

statement which describes the world as it is

Introduction to Demand and Supply

class="introduction"

Farmer's Market

Organic vegetables and fruits that are grown and sold within a specific geographical region should, in theory, cost less than conventional produce because the transportation costs are less.

That is not, however, usually the case. (Credit: modification of work by Natalie Maynor/Flickr
r Creative Commons)

**Note:****Why Can We Not Get Enough of Organic?**

Organic food is increasingly popular, not just in the United States, but worldwide. At one time, consumers had to go to specialty stores or farmer's markets to find organic produce. Now it is available in most grocery stores. In short, organic is part of the mainstream.

Ever wonder why organic food costs more than conventional food? Why, say, does an organic Fuji apple cost \$1.99 a pound, while its conventional counterpart costs \$1.49 a pound? The same price relationship is true for just about every organic product on the market. If many organic foods are locally grown, would they not take less time to get to market and therefore be cheaper? What are the forces that keep those prices from coming down? Turns out those forces have a lot to do with this chapter's topic: demand and supply.

Note:**Introduction to Demand and Supply**

In this chapter, you will learn about:

- Demand, Supply, and Equilibrium in Markets for Goods and Services
- Shifts in Demand and Supply for Goods and Services
- Changes in Equilibrium Price and Quantity: The Four-Step Process
- Price Ceilings and Price Floors

An auction bidder pays thousands of dollars for a dress Whitney Houston wore. A collector spends a small fortune for a few drawings by John Lennon. People usually react to purchases like these in two ways: their jaw drops because they think these are high prices to pay for such goods or they think these are rare, desirable items and the amount paid seems right.

Note:

Visit this [website](#) to read a list of bizarre items that have been purchased for their ties to celebrities. These examples represent an interesting facet of demand and supply.



When economists talk about prices, they are less interested in making judgments than in gaining a practical understanding of what determines prices and why prices change. Consider a price most of us contend with weekly: that of a gallon of gas. Why was the average price of gasoline in the United States \$3.71 per gallon in June 2014? Why did the price for gasoline fall sharply to \$2.07 per gallon by January 2015? To explain these price movements, economists focus on the determinants of what gasoline buyers are willing to pay and what gasoline sellers are willing to accept.

As it turns out, the price of gasoline in June of any given year is nearly always higher than the price in January of that same year; over recent decades, gasoline prices in midsummer have averaged about 10 cents per gallon more than their midwinter low. The likely reason is that people drive more in the summer, and are also willing to pay more for gas, but that does not explain how steeply gas prices fell. Other factors were at work during those six months, such as increases in supply and decreases in the demand for crude oil.

This chapter introduces the economic model of demand and supply—one of the most powerful models in all of economics. The discussion here begins by examining how demand and supply determine the price and the quantity sold in markets for goods and services, and how changes in demand and supply lead to changes in prices and quantities.

Demand, Supply, and Equilibrium in Markets for Goods and Services

By the end of this section, you will be able to:

- Explain demand, quantity demanded, and the law of demand
- Identify a demand curve and a supply curve
- Explain supply, quantity supply, and the law of supply
- Explain equilibrium, equilibrium price, and equilibrium quantity

First let's first focus on what economists mean by demand, what they mean by supply, and then how demand and supply interact in a market.

Demand for Goods and Services

Economists use the term **demand** to refer to the amount of some good or service consumers are willing and able to purchase at each price. Demand is based on needs and wants—a consumer may be able to differentiate between a need and a want, but from an economist's perspective they are the same thing. Demand is also based on ability to pay. If you cannot pay for it, you have no effective demand.

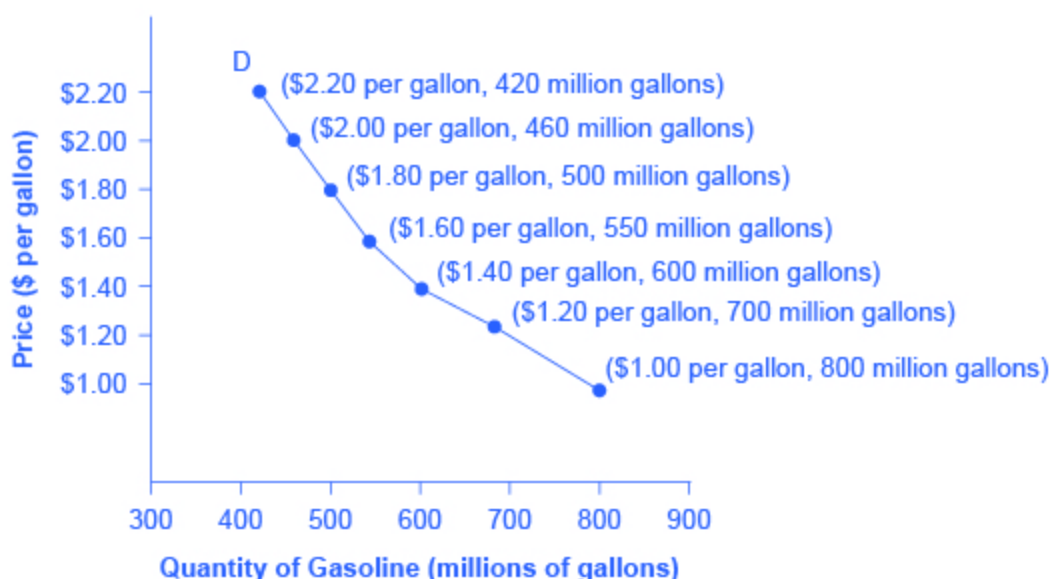
What a buyer pays for a unit of the specific good or service is called **price**. The total number of units purchased at that price is called the **quantity demanded**. A rise in price of a good or service almost always decreases the quantity demanded of that good or service. Conversely, a fall in price will increase the quantity demanded. When the price of a gallon of gasoline goes up, for example, people look for ways to reduce their consumption by combining several errands, commuting by carpool or mass transit, or taking weekend or vacation trips closer to home. Economists call this inverse relationship between price and quantity demanded the **law of demand**. The law of demand assumes that all other variables that affect demand (to be explained in the next module) are held constant.

An example from the market for gasoline can be shown in the form of a table or a graph. A table that shows the quantity demanded at each price, such as [\[link\]](#), is called a **demand schedule**. Price in this case is measured in dollars per gallon of gasoline. The quantity demanded is measured in

millions of gallons over some time period (for example, per day or per year) and over some geographic area (like a state or a country). A **demand curve** shows the relationship between price and quantity demanded on a graph like [\[link\]](#), with quantity on the horizontal axis and the price per gallon on the vertical axis. (Note that this is an exception to the normal rule in mathematics that the independent variable (x) goes on the horizontal axis and the dependent variable (y) goes on the vertical. Economics is not math.)

The demand schedule shown by [\[link\]](#) and the demand curve shown by the graph in [\[link\]](#) are two ways of describing the same relationship between price and quantity demanded.

A Demand Curve for Gasoline



The demand schedule shows that as price rises, quantity demanded decreases, and vice versa. These points are then graphed, and the line connecting them is the demand curve (D). The downward slope of the demand curve again illustrates the law of demand—the inverse relationship between prices and quantity demanded.

Price (per gallon)	Quantity Demanded (millions of gallons)
\$1.00	800
\$1.20	700
\$1.40	600
\$1.60	550
\$1.80	500
\$2.00	460
\$2.20	420

Price and Quantity Demanded of Gasoline

Demand curves will appear somewhat different for each product. They may appear relatively steep or flat, or they may be straight or curved. Nearly all demand curves share the fundamental similarity that they slope down from left to right. So demand curves embody the law of demand: As the price increases, the quantity demanded decreases, and conversely, as the price decreases, the quantity demanded increases.

Confused about these different types of demand? Read the next Clear It Up feature.

Note:

Is demand the same as quantity demanded?

In economic terminology, demand is not the same as quantity demanded. When economists talk about demand, they mean the relationship between a range of prices and the quantities demanded at those prices, as illustrated by a demand curve or a demand schedule. When economists talk about quantity demanded, they mean only a certain point on the demand curve, or

one quantity on the demand schedule. In short, demand refers to the curve and quantity demanded refers to the (specific) point on the curve.

Supply of Goods and Services

When economists talk about **supply**, they mean the amount of some good or service a producer is willing to supply at each price. Price is what the producer receives for selling one unit of a good or service. A rise in price almost always leads to an increase in the **quantity supplied** of that good or service, while a fall in price will decrease the quantity supplied. When the price of gasoline rises, for example, it encourages profit-seeking firms to take several actions: expand exploration for oil reserves; drill for more oil; invest in more pipelines and oil tankers to bring the oil to plants where it can be refined into gasoline; build new oil refineries; purchase additional pipelines and trucks to ship the gasoline to gas stations; and open more gas stations or keep existing gas stations open longer hours. Economists call this positive relationship between price and quantity supplied—that a higher price leads to a higher quantity supplied and a lower price leads to a lower quantity supplied—the **law of supply**. The law of supply assumes that all other variables that affect supply (to be explained in the next module) are held constant.

Still unsure about the different types of supply? See the following Clear It Up feature.

Note:

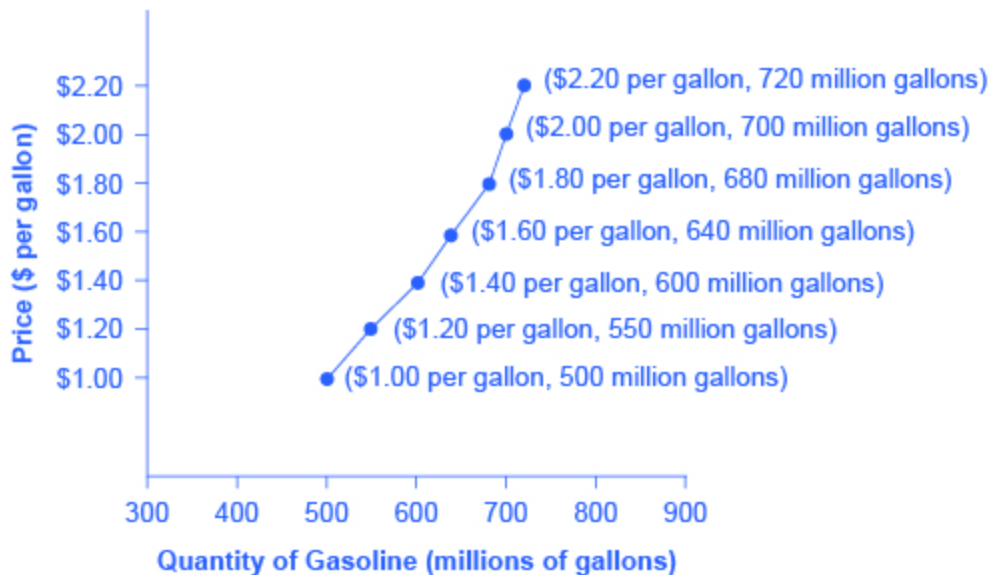
Is supply the same as quantity supplied?

In economic terminology, supply is not the same as quantity supplied. When economists refer to supply, they mean the relationship between a range of prices and the quantities supplied at those prices, a relationship that can be illustrated with a supply curve or a supply schedule. When economists refer to quantity supplied, they mean only a certain point on the supply curve, or one quantity on the supply schedule. In short, supply

refers to the curve and quantity supplied refers to the (specific) point on the curve.

[\[link\]](#) illustrates the law of supply, again using the market for gasoline as an example. Like demand, supply can be illustrated using a table or a graph. A **supply schedule** is a table, like [\[link\]](#), that shows the quantity supplied at a range of different prices. Again, price is measured in dollars per gallon of gasoline and quantity supplied is measured in millions of gallons. A **supply curve** is a graphic illustration of the relationship between price, shown on the vertical axis, and quantity, shown on the horizontal axis. The supply schedule and the supply curve are just two different ways of showing the same information. Notice that the horizontal and vertical axes on the graph for the supply curve are the same as for the demand curve.

A Supply Curve for Gasoline



The supply schedule is the table that shows quantity supplied of gasoline at each price. As price rises, quantity supplied also increases, and vice versa. The supply curve (S) is created by graphing the points from the supply schedule and then connecting them. The upward slope of the supply curve illustrates the law of supply—that a higher price leads to a higher quantity supplied, and vice versa.

Price (per gallon)	Quantity Supplied (millions of gallons)
\$1.00	500
\$1.20	550
\$1.40	600
\$1.60	640
\$1.80	680
\$2.00	700
\$2.20	720

Price and Supply of Gasoline

The shape of supply curves will vary somewhat according to the product: steeper, flatter, straighter, or curved. Nearly all supply curves, however, share a basic similarity: they slope up from left to right and illustrate the law of supply: as the price rises, say, from \$1.00 per gallon to \$2.20 per gallon, the quantity supplied increases from 500 gallons to 720 gallons. Conversely, as the price falls, the quantity supplied decreases.

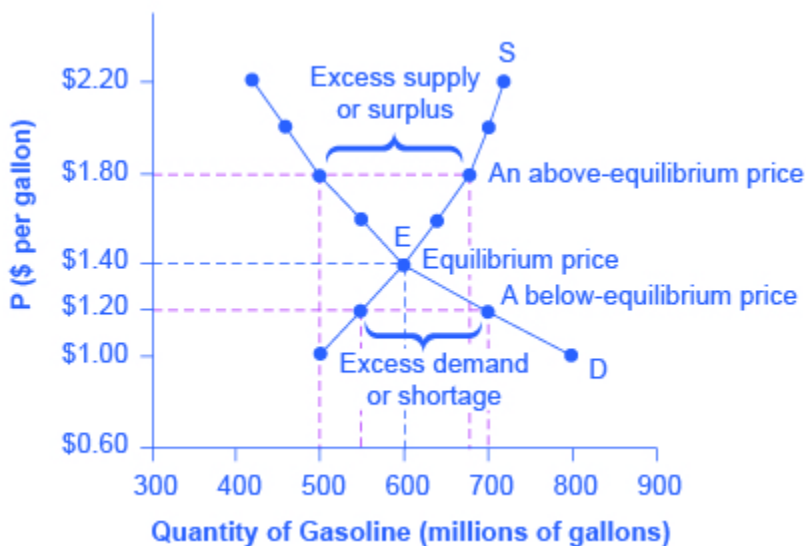
Equilibrium—Where Demand and Supply Intersect

Because the graphs for demand and supply curves both have price on the vertical axis and quantity on the horizontal axis, the demand curve and supply curve for a particular good or service can appear on the same graph.

Together, demand and supply determine the price and the quantity that will be bought and sold in a market.

[\[link\]](#) illustrates the interaction of demand and supply in the market for gasoline. The demand curve (D) is identical to [\[link\]](#). The supply curve (S) is identical to [\[link\]](#). [\[link\]](#) contains the same information in tabular form.

Demand and Supply for Gasoline



The demand curve (D) and the supply curve (S) intersect at the equilibrium point E, with a price of \$1.40 and a quantity of 600. The equilibrium is the only price where quantity demanded is equal to quantity supplied. At a price above equilibrium like \$1.80, quantity supplied exceeds the quantity demanded, so there is excess supply. At a price below equilibrium such as \$1.20, quantity demanded exceeds quantity supplied, so there is excess demand.

Price (per gallon)	Quantity demanded (millions of gallons)	Quantity supplied (millions of gallons)
\$1.00	800	500
\$1.20	700	550
\$1.40	600	600
\$1.60	550	640
\$1.80	500	680
\$2.00	460	700
\$2.20	420	720

Price, Quantity Demanded, and Quantity Supplied

Remember this: When two lines on a diagram cross, this intersection usually means something. The point where the supply curve (S) and the demand curve (D) cross, designated by point E in [\[link\]](#), is called the **equilibrium**. The **equilibrium price** is the only price where the plans of consumers and the plans of producers agree—that is, where the amount of the product consumers want to buy (quantity demanded) is equal to the amount producers want to sell (quantity supplied). This common quantity is called the **equilibrium quantity**. At any other price, the quantity demanded does not equal the quantity supplied, so the market is not in equilibrium at that price.

In [\[link\]](#), the equilibrium price is \$1.40 per gallon of gasoline and the equilibrium quantity is 600 million gallons. If you had only the demand and supply schedules, and not the graph, you could find the equilibrium by looking for the price level on the tables where the quantity demanded and the quantity supplied are equal.

The word “equilibrium” means “balance.” If a market is at its equilibrium price and quantity, then it has no reason to move away from that point. However, if a market is not at equilibrium, then economic pressures arise to move the market toward the equilibrium price and the equilibrium quantity.

Imagine, for example, that the price of a gallon of gasoline was above the equilibrium price—that is, instead of \$1.40 per gallon, the price is \$1.80 per gallon. This above-equilibrium price is illustrated by the dashed horizontal line at the price of \$1.80 in [\[link\]](#). At this higher price, the quantity demanded drops from 600 to 500. This decline in quantity reflects how consumers react to the higher price by finding ways to use less gasoline.

Moreover, at this higher price of \$1.80, the quantity of gasoline supplied rises from the 600 to 680, as the higher price makes it more profitable for gasoline producers to expand their output. Now, consider how quantity demanded and quantity supplied are related at this above-equilibrium price. Quantity demanded has fallen to 500 gallons, while quantity supplied has risen to 680 gallons. In fact, at any above-equilibrium price, the quantity supplied exceeds the quantity demanded. We call this an **excess supply** or a **surplus**.

With a surplus, gasoline accumulates at gas stations, in tanker trucks, in pipelines, and at oil refineries. This accumulation puts pressure on gasoline sellers. If a surplus remains unsold, those firms involved in making and selling gasoline are not receiving enough cash to pay their workers and to cover their expenses. In this situation, some producers and sellers will want to cut prices, because it is better to sell at a lower price than not to sell at all. Once some sellers start cutting prices, others will follow to avoid losing sales. These price reductions in turn will stimulate a higher quantity demanded. So, if the price is above the equilibrium level, incentives built into the structure of demand and supply will create pressures for the price to fall toward the equilibrium.

Now suppose that the price is below its equilibrium level at \$1.20 per gallon, as the dashed horizontal line at this price in [\[link\]](#) shows. At this lower price, the quantity demanded increases from 600 to 700 as drivers take longer trips, spend more minutes warming up the car in the driveway in wintertime, stop sharing rides to work, and buy larger cars that get fewer

miles to the gallon. However, the below-equilibrium price reduces gasoline producers' incentives to produce and sell gasoline, and the quantity supplied falls from 600 to 550.

When the price is below equilibrium, there is **excess demand**, or a **shortage**—that is, at the given price the quantity demanded, which has been stimulated by the lower price, now exceeds the quantity supplied, which had been depressed by the lower price. In this situation, eager gasoline buyers mob the gas stations, only to find many stations running short of fuel. Oil companies and gas stations recognize that they have an opportunity to make higher profits by selling what gasoline they have at a higher price. As a result, the price rises toward the equilibrium level. Read [Demand, Supply, and Efficiency](#) for more discussion on the importance of the demand and supply model.

Key Concepts and Summary

A demand schedule is a table that shows the quantity demanded at different prices in the market. A demand curve shows the relationship between quantity demanded and price in a given market on a graph. The law of demand states that a higher price typically leads to a lower quantity demanded.

A supply schedule is a table that shows the quantity supplied at different prices in the market. A supply curve shows the relationship between quantity supplied and price on a graph. The law of supply says that a higher price typically leads to a higher quantity supplied.

The equilibrium price and equilibrium quantity occur where the supply and demand curves cross. The equilibrium occurs where the quantity demanded is equal to the quantity supplied. If the price is below the equilibrium level, then the quantity demanded will exceed the quantity supplied. Excess demand or a shortage will exist. If the price is above the equilibrium level, then the quantity supplied will exceed the quantity demanded. Excess supply or a surplus will exist. In either case, economic pressures will push the price toward the equilibrium level.

Self-Check Question

Exercise:

Problem:

Review [\[link\]](#). Suppose the price of gasoline is \$1.60 per gallon. Is the quantity demanded higher or lower than at the equilibrium price of \$1.40 per gallon? And what about the quantity supplied? Is there a shortage or a surplus in the market? If so, of how much?

Solution:

Since \$1.60 per gallon is above the equilibrium price, the quantity demanded would be lower at 550 gallons and the quantity supplied would be higher at 640 gallons. (These results are due to the laws of demand and supply, respectively.) The outcome of lower Q_d and higher Q_s would be a surplus in the gasoline market of $640 - 550 = 90$ gallons.

Review Questions

Exercise:

Problem: What determines the level of prices in a market?

Exercise:

Problem:

What does a downward-sloping demand curve mean about how buyers in a market will react to a higher price?

Exercise:

Problem:

Will demand curves have the same exact shape in all markets? If not, how will they differ?

Exercise:

Problem:

Will supply curves have the same shape in all markets? If not, how will they differ?

Exercise:

Problem:

What is the relationship between quantity demanded and quantity supplied at equilibrium? What is the relationship when there is a shortage? What is the relationship when there is a surplus?

Exercise:

Problem:

How can you locate the equilibrium point on a demand and supply graph?

Exercise:

Problem:

If the price is above the equilibrium level, would you predict a surplus or a shortage? If the price is below the equilibrium level, would you predict a surplus or a shortage? Why?

Exercise:

Problem:

When the price is above the equilibrium, explain how market forces move the market price to equilibrium. Do the same when the price is below the equilibrium.

Exercise:

Problem:

What is the difference between the demand and the quantity demanded of a product, say milk? Explain in words and show the difference on a graph with a demand curve for milk.

Exercise:**Problem:**

What is the difference between the supply and the quantity supplied of a product, say milk? Explain in words and show the difference on a graph with the supply curve for milk.

Critical Thinking Questions**Exercise:****Problem:**

Review [\[link\]](#). Suppose the government decided that, since gasoline is a necessity, its price should be legally capped at \$1.30 per gallon. What do you anticipate would be the outcome in the gasoline market?

Exercise:**Problem:**

Explain why the following statement is false: “In the goods market, no buyer would be willing to pay more than the equilibrium price.”

Exercise:**Problem:**

Explain why the following statement is false: “In the goods market, no seller would be willing to sell for less than the equilibrium price.”

Problems

Exercise:

Problem:

Review [\[link\]](#) again. Suppose the price of gasoline is \$1.00. Will the quantity demanded be lower or higher than at the equilibrium price of \$1.40 per gallon? Will the quantity supplied be lower or higher? Is there a shortage or a surplus in the market? If so, of how much?

References

Costanza, Robert, and Lisa Wainger. "No Accounting For Nature: How Conventional Economics Distorts the Value of Things." *The Washington Post*. September 2, 1990.

European Commission: Agriculture and Rural Development. 2013. "Overview of the CAP Reform: 2014-2024." Accessed April 13, 2015. <http://ec.europa.eu/agriculture/cap-post-2013/>.

Radford, R. A. "The Economic Organisation of a P.O.W. Camp." *Economica*. no. 48 (1945): 189-201. <http://www.jstor.org/stable/2550133>.

Glossary

demand curve

a graphic representation of the relationship between price and quantity demanded of a certain good or service, with quantity on the horizontal axis and the price on the vertical axis

demand schedule

a table that shows a range of prices for a certain good or service and the quantity demanded at each price

demand

the relationship between price and the quantity demanded of a certain good or service

equilibrium price

the price where quantity demanded is equal to quantity supplied

equilibrium quantity

the quantity at which quantity demanded and quantity supplied are equal for a certain price level

equilibrium

the situation where quantity demanded is equal to the quantity supplied; the combination of price and quantity where there is no economic pressure from surpluses or shortages that would cause price or quantity to change

excess demand

at the existing price, the quantity demanded exceeds the quantity supplied; also called a shortage

excess supply

at the existing price, quantity supplied exceeds the quantity demanded; also called a surplus

law of demand

the common relationship that a higher price leads to a lower quantity demanded of a certain good or service and a lower price leads to a higher quantity demanded, while all other variables are held constant

law of supply

the common relationship that a higher price leads to a greater quantity supplied and a lower price leads to a lower quantity supplied, while all other variables are held constant

price

what a buyer pays for a unit of the specific good or service

quantity demanded

the total number of units of a good or service consumers are willing to purchase at a given price

quantity supplied

the total number of units of a good or service producers are willing to sell at a given price

shortage

at the existing price, the quantity demanded exceeds the quantity supplied; also called excess demand

supply curve

a line that shows the relationship between price and quantity supplied on a graph, with quantity supplied on the horizontal axis and price on the vertical axis

supply schedule

a table that shows a range of prices for a good or service and the quantity supplied at each price

supply

the relationship between price and the quantity supplied of a certain good or service

surplus

at the existing price, quantity supplied exceeds the quantity demanded; also called excess supply

Shifts in Demand and Supply for Goods and Services

By the end of this section, you will be able to:

- Identify factors that affect demand
- Graph demand curves and demand shifts
- Identify factors that affect supply
- Graph supply curves and supply shifts

The previous module explored how price affects the quantity demanded and the quantity supplied. The result was the demand curve and the supply curve. Price, however, is not the only thing that influences demand. Nor is it the only thing that influences supply. For example, how is demand for vegetarian food affected if, say, health concerns cause more consumers to avoid eating meat? Or how is the supply of diamonds affected if diamond producers discover several new diamond mines? What are the major factors, in addition to the price, that influence demand or supply?

What Factors Affect Demand?

We defined demand as the amount of some product a consumer is willing and able to purchase at each price. That suggests at least two factors in addition to price that affect demand. Willingness to purchase suggests a desire, based on what economists call tastes and preferences. If you neither need nor want something, you will not buy it. Ability to purchase suggests that income is important. Professors are usually able to afford better housing and transportation than students, because they have more income. Prices of related goods can affect demand also. If you need a new car, the price of a Honda may affect your demand for a Ford. Finally, the size or composition of the population can affect demand. The more children a family has, the greater their demand for clothing. The more driving-age children a family has, the greater their demand for car insurance, and the less for diapers and baby formula.

These factors matter both for demand by an individual and demand by the market as a whole. Exactly how do these various factors affect demand, and

how do we show the effects graphically? To answer those questions, we need the *ceteris paribus* assumption.

The *Ceteris Paribus* Assumption

A demand curve or a supply curve is a relationship between two, and only two, variables: quantity on the horizontal axis and price on the vertical axis. The assumption behind a demand curve or a supply curve is that no relevant economic factors, other than the product's price, are changing. Economists call this assumption **ceteris paribus**, a Latin phrase meaning “other things being equal.” Any given demand or supply curve is based on the *ceteris paribus* assumption that all else is held equal. A demand curve or a supply curve is a relationship between two, and only two, variables when all other variables are kept constant. If all else is not held equal, then the laws of supply and demand will not necessarily hold, as the following Clear It Up feature shows.

Note:

When does *ceteris paribus* apply?

Ceteris paribus is typically applied when we look at how changes in price affect demand or supply, but *ceteris paribus* can be applied more generally. In the real world, demand and supply depend on more factors than just price. For example, a consumer's demand depends on income and a producer's supply depends on the cost of producing the product. How can we analyze the effect on demand or supply if multiple factors are changing at the same time—say price rises and income falls? The answer is that we examine the changes one at a time, assuming the other factors are held constant.

For example, we can say that an increase in the price reduces the amount consumers will buy (assuming income, and anything else that affects demand, is unchanged). Additionally, a decrease in income reduces the amount consumers can afford to buy (assuming price, and anything else that affects demand, is unchanged). This is what the *ceteris paribus* assumption really means. In this particular case, after we analyze each factor separately, we can combine the results. The amount consumers buy

falls for two reasons: first because of the higher price and second because of the lower income.

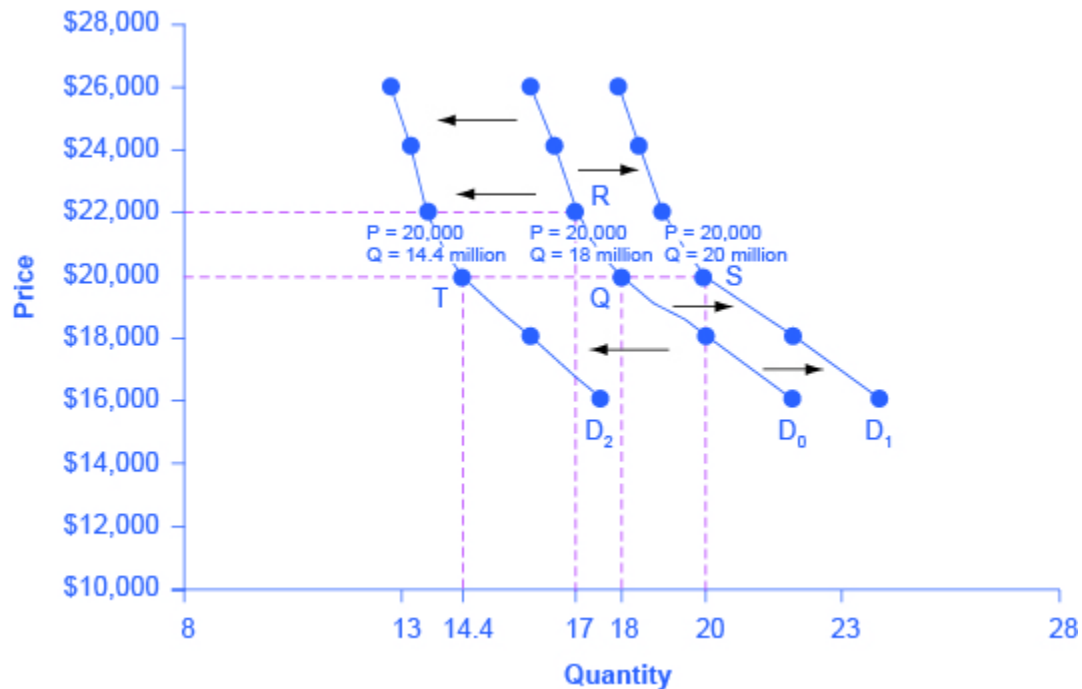
How Does Income Affect Demand?

Let's use income as an example of how factors other than price affect demand. [\[link\]](#) shows the initial demand for automobiles as D_0 . At point Q, for example, if the price is \$20,000 per car, the quantity of cars demanded is 18 million. D_0 also shows how the quantity of cars demanded would change as a result of a higher or lower price. For example, if the price of a car rose to \$22,000, the quantity demanded would decrease to 17 million, at point R.

The original demand curve D_0 , like every demand curve, is based on the *ceteris paribus* assumption that no other economically relevant factors change. Now imagine that the economy expands in a way that raises the incomes of many people, making cars more affordable. How will this affect demand? How can we show this graphically?

Return to [\[link\]](#). The price of cars is still \$20,000, but with higher incomes, the quantity demanded has now increased to 20 million cars, shown at point S. As a result of the higher income levels, the demand curve shifts to the right to the new demand curve D_1 , indicating an increase in demand. [\[link\]](#) shows clearly that this increased demand would occur at every price, not just the original one.

Shifts in Demand: A Car Example



Increased demand means that at every given price, the quantity demanded is higher, so that the demand curve shifts to the right from D_0 to D_1 . Decreased demand means that at every given price, the quantity demanded is lower, so that the demand curve shifts to the left from D_0 to D_2 .

Price	Decrease to D_2	Original Quantity Demanded D_0	Increase to D_1
\$16,000	17.6 million	22.0 million	24.0 million
\$18,000	16.0 million	20.0 million	22.0 million

Price	Decrease to D_2	Original Quantity Demanded D_0	Increase to D_1
\$20,000	14.4 million	18.0 million	20.0 million
\$22,000	13.6 million	17.0 million	19.0 million
\$24,000	13.2 million	16.5 million	18.5 million
\$26,000	12.8 million	16.0 million	18.0 million

Price and Demand Shifts: A Car Example

Now, imagine that the economy slows down so that many people lose their jobs or work fewer hours, reducing their incomes. In this case, the decrease in income would lead to a lower quantity of cars demanded at every given price, and the original demand curve D_0 would shift left to D_2 . The shift from D_0 to D_2 represents such a decrease in demand: At any given price level, the quantity demanded is now lower. In this example, a price of \$20,000 means 18 million cars sold along the original demand curve, but only 14.4 million sold after demand fell.

When a demand curve shifts, it does not mean that the quantity demanded by every individual buyer changes by the same amount. In this example, not everyone would have higher or lower income and not everyone would buy or not buy an additional car. Instead, a shift in a demand curve captures an pattern for the market as a whole.

In the previous section, we argued that higher income causes greater demand at every price. This is true for most goods and services. For some—luxury cars, vacations in Europe, and fine jewelry—the effect of a rise in income can be especially pronounced. A product whose demand rises when income rises, and vice versa, is called a **normal good**. A few exceptions to

this pattern do exist. As incomes rise, many people will buy fewer generic brand groceries and more name brand groceries. They are less likely to buy used cars and more likely to buy new cars. They will be less likely to rent an apartment and more likely to own a home, and so on. A product whose demand falls when income rises, and vice versa, is called an **inferior good**. In other words, when income increases, the demand curve shifts to the left.

Other Factors That Shift Demand Curves

Income is not the only factor that causes a shift in demand. Other things that change demand include tastes and preferences, the composition or size of the population, the prices of related goods, and even expectations. A change in any one of the underlying factors that determine what quantity people are willing to buy at a given price will cause a shift in demand. Graphically, the new demand curve lies either to the right (an increase) or to the left (a decrease) of the original demand curve. Let's look at these factors.

Changing Tastes or Preferences

From 1980 to 2014, the per-person consumption of chicken by Americans rose from 48 pounds per year to 85 pounds per year, and consumption of beef fell from 77 pounds per year to 54 pounds per year, according to the U.S. Department of Agriculture (USDA). Changes like these are largely due to movements in taste, which change the quantity of a good demanded at every price: that is, they shift the demand curve for that good, rightward for chicken and leftward for beef.

Changes in the Composition of the Population

The proportion of elderly citizens in the United States population is rising. It rose from 9.8% in 1970 to 12.6% in 2000, and will be a projected (by the U.S. Census Bureau) 20% of the population by 2030. A society with relatively more children, like the United States in the 1960s, will have greater demand for goods and services like tricycles and day care facilities. A society with relatively more elderly persons, as the United States is projected to have by 2030, has a higher demand for nursing homes and hearing aids. Similarly, changes in the size of the population can affect the

demand for housing and many other goods. Each of these changes in demand will be shown as a shift in the demand curve.

The demand for a product can also be affected by changes in the prices of related goods such as substitutes or complements. A **substitute** is a good or service that can be used in place of another good or service. As electronic books, like this one, become more available, you would expect to see a decrease in demand for traditional printed books. A lower price for a substitute decreases demand for the other product. For example, in recent years as the price of tablet computers has fallen, the quantity demanded has increased (because of the law of demand). Since people are purchasing tablets, there has been a decrease in demand for laptops, which can be shown graphically as a leftward shift in the demand curve for laptops. A higher price for a substitute good has the reverse effect.

Other goods are **complements** for each other, meaning that the goods are often used together, because consumption of one good tends to enhance consumption of the other. Examples include breakfast cereal and milk; notebooks and pens or pencils, golf balls and golf clubs; gasoline and sport utility vehicles; and the five-way combination of bacon, lettuce, tomato, mayonnaise, and bread. If the price of golf clubs rises, since the quantity demanded of golf clubs falls (because of the law of demand), demand for a complement good like golf balls decreases, too. Similarly, a higher price for skis would shift the demand curve for a complement good like ski resort trips to the left, while a lower price for a complement has the reverse effect.

Changes in Expectations about Future Prices or Other Factors that Affect Demand

While it is clear that the price of a good affects the quantity demanded, it is also true that expectations about the future price (or expectations about tastes and preferences, income, and so on) can affect demand. For example, if people hear that a hurricane is coming, they may rush to the store to buy flashlight batteries and bottled water. If people learn that the price of a good like coffee is likely to rise in the future, they may head for the store to stock up on coffee now. These changes in demand are shown as shifts in the curve. Therefore, a **shift in demand** happens when a change in some economic factor (other than price) causes a different quantity to be

demanded at every price. The following Work It Out feature shows how this happens.

Note:

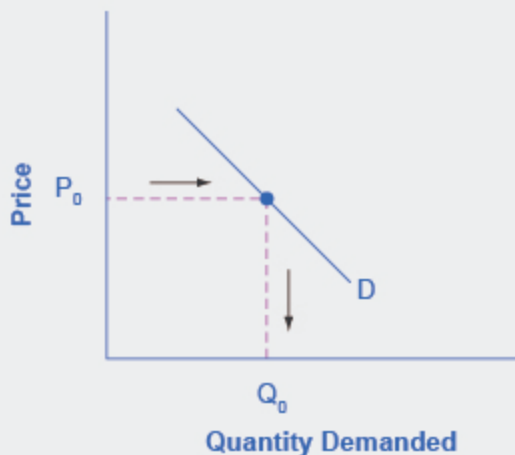
Shift in Demand

A shift in demand means that at any price (and at every price), the quantity demanded will be different than it was before. Following is an example of a shift in demand due to an income increase.

Step 1. Draw the graph of a demand curve for a normal good like pizza.

Pick a price (like P_0). Identify the corresponding Q_0 . An example is shown in [\[link\]](#).

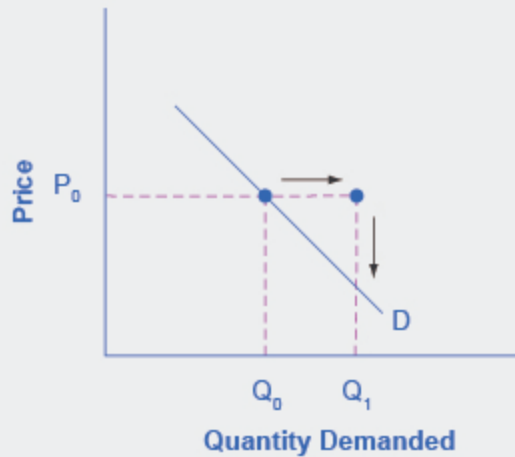
Demand Curve



The demand curve can be used to identify how much consumers would buy at any given price.

Step 2. Suppose income increases. As a result of the change, are consumers going to buy more or less pizza? The answer is more. Draw a dotted horizontal line from the chosen price, through the original quantity demanded, to the new point with the new Q_1 . Draw a dotted vertical line down to the horizontal axis and label the new Q_1 . An example is provided in [\[link\]](#).

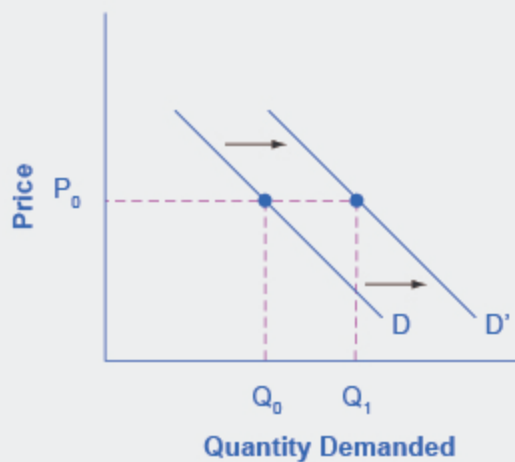
Demand Curve with Income Increase



With an increase in income, consumers will purchase larger quantities, pushing demand to the right.

Step 3. Now, shift the curve through the new point. You will see that an increase in income causes an upward (or rightward) shift in the demand curve, so that at any price the quantities demanded will be higher, as shown in [\[link\]](#).

Demand Curve Shifted Right

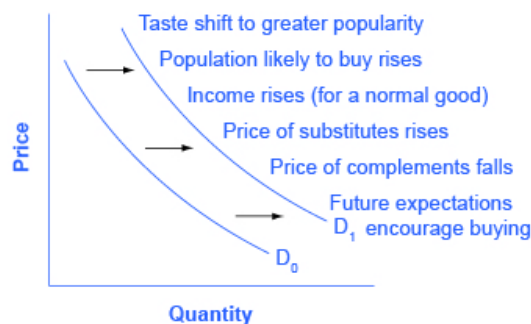


With an increase in income, consumers will purchase larger quantities, pushing demand to the right, and causing the demand curve to shift right.

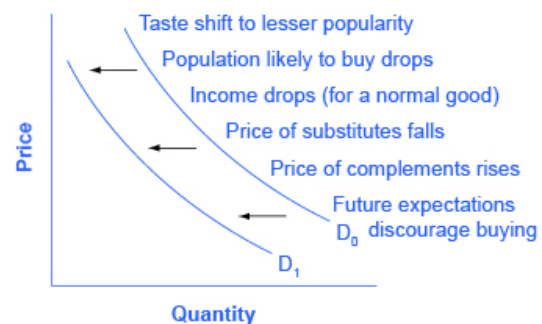
Summing Up Factors That Change Demand

Six factors that can shift demand curves are summarized in [\[link\]](#). The direction of the arrows indicates whether the demand curve shifts represent an increase in demand or a decrease in demand. Notice that a change in the price of the good or service itself is not listed among the factors that can shift a demand curve. A change in the price of a good or service causes a movement along a specific demand curve, and it typically leads to some change in the quantity demanded, but it does not shift the demand curve.

Factors That Shift Demand Curves



(a) Factors that increase demand



(b) Factors that decrease demand

- (a) A list of factors that can cause an increase in demand from D_0 to D_1 . (b) The same factors, if their direction is reversed, can cause a decrease in demand from D_0 to D_1 .

When a demand curve shifts, it will then intersect with a given supply curve at a different equilibrium price and quantity. We are, however, getting ahead of our story. Before discussing how changes in demand can affect equilibrium price and quantity, we first need to discuss shifts in supply curves.

How Production Costs Affect Supply

A supply curve shows how quantity supplied will change as the price rises and falls, assuming *ceteris paribus* so that no other economically relevant factors are changing. If other factors relevant to supply do change, then the entire supply curve will shift. Just as a shift in demand is represented by a change in the quantity demanded at every price, a **shift in supply** means a change in the quantity supplied at every price.

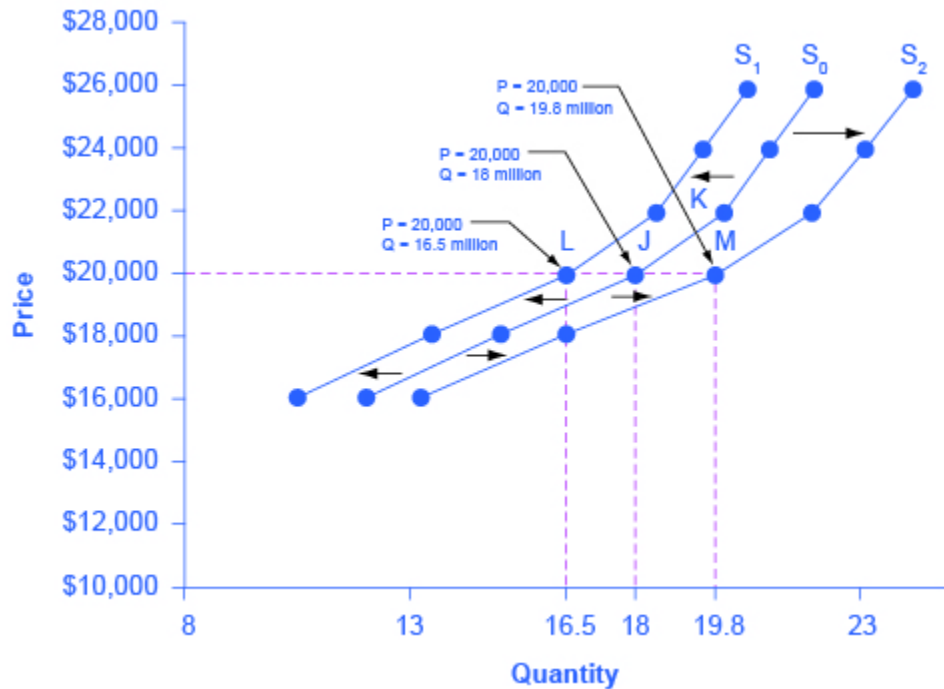
In thinking about the factors that affect supply, remember what motivates firms: profits, which are the difference between revenues and costs. Goods and services are produced using combinations of labor, materials, and machinery, or what we call **inputs** or **factors of production**. If a firm faces lower costs of production, while the prices for the good or service the firm produces remain unchanged, a firm's profits go up. When a firm's profits increase, it is more motivated to produce output, since the more it produces the more profit it will earn. So, when costs of production fall, a firm will tend to supply a larger quantity at any given price for its output. This can be shown by the supply curve shifting to the right.

Take, for example, a messenger company that delivers packages around a city. The company may find that buying gasoline is one of its main costs. If the price of gasoline falls, then the company will find it can deliver messages more cheaply than before. Since lower costs correspond to higher profits, the messenger company may now supply more of its services at any given price. For example, given the lower gasoline prices, the company can now serve a greater area, and increase its supply.

Conversely, if a firm faces higher costs of production, then it will earn lower profits at any given selling price for its products. As a result, a higher cost of production typically causes a firm to supply a smaller quantity at any given price. In this case, the supply curve shifts to the left.

Consider the supply for cars, shown by curve S_0 in [\[link\]](#). Point J indicates that if the price is \$20,000, the quantity supplied will be 18 million cars. If the price rises to \$22,000 per car, *ceteris paribus*, the quantity supplied will rise to 20 million cars, as point K on the S_0 curve shows. The same information can be shown in table form, as in [\[link\]](#).

Shifts in Supply: A Car Example



Decreased supply means that at every given price, the quantity supplied is lower, so that the supply curve shifts to the left, from S_0 to S_1 . Increased supply means that at every given price, the quantity supplied is higher, so that the supply curve shifts to the right, from S_0 to S_2 .

Price	Decrease to S_1	Original Quantity Supplied S_0	Increase to S_2
\$16,000	10.5 million	12.0 million	13.2 million

Price	Decrease to S_1	Original Quantity Supplied S_0	Increase to S_2
\$18,000	13.5 million	15.0 million	16.5 million
\$20,000	16.5 million	18.0 million	19.8 million
\$22,000	18.5 million	20.0 million	22.0 million
\$24,000	19.5 million	21.0 million	23.1 million
\$26,000	20.5 million	22.0 million	24.2 million

Price and Shifts in Supply: A Car Example

Now, imagine that the price of steel, an important ingredient in manufacturing cars, rises, so that producing a car has become more expensive. At any given price for selling cars, car manufacturers will react by supplying a lower quantity. This can be shown graphically as a leftward shift of supply, from S_0 to S_1 , which indicates that at any given price, the quantity supplied decreases. In this example, at a price of \$20,000, the quantity supplied decreases from 18 million on the original supply curve (S_0) to 16.5 million on the supply curve S_1 , which is labeled as point L.

Conversely, if the price of steel decreases, producing a car becomes less expensive. At any given price for selling cars, car manufacturers can now expect to earn higher profits, so they will supply a higher quantity. The shift of supply to the right, from S_0 to S_2 , means that at all prices, the quantity supplied has increased. In this example, at a price of \$20,000, the quantity supplied increases from 18 million on the original supply curve (S_0) to 19.8 million on the supply curve S_2 , which is labeled M.

Other Factors That Affect Supply

In the example above, we saw that changes in the prices of inputs in the production process will affect the cost of production and thus the supply. Several other things affect the cost of production, too, such as changes in weather or other natural conditions, new technologies for production, and some government policies.

The cost of production for many agricultural products will be affected by changes in natural conditions. For example, in 2014 the Manchurian Plain in Northeastern China, which produces most of the country's wheat, corn, and soybeans, experienced its most severe drought in 50 years. A drought decreases the supply of agricultural products, which means that at any given price, a lower quantity will be supplied; conversely, especially good weather would shift the supply curve to the right.

When a firm discovers a new technology that allows the firm to produce at a lower cost, the supply curve will shift to the right, as well. For instance, in the 1960s a major scientific effort nicknamed the Green Revolution focused on breeding improved seeds for basic crops like wheat and rice. By the early 1990s, more than two-thirds of the wheat and rice in low-income countries around the world was grown with these Green Revolution seeds—and the harvest was twice as high per acre. A technological improvement that reduces costs of production will shift supply to the right, so that a greater quantity will be produced at any given price.

Government policies can affect the cost of production and the supply curve through taxes, regulations, and subsidies. For example, the U.S. government imposes a tax on alcoholic beverages that collects about \$8 billion per year from producers. Taxes are treated as costs by businesses. Higher costs decrease supply for the reasons discussed above. Other examples of policy that can affect cost are the wide array of government regulations that require firms to spend money to provide a cleaner environment or a safer workplace; complying with regulations increases costs.

A government subsidy, on the other hand, is the opposite of a tax. A subsidy occurs when the government pays a firm directly or reduces the firm's taxes if the firm carries out certain actions. From the firm's perspective, taxes or

regulations are an additional cost of production that shifts supply to the left, leading the firm to produce a lower quantity at every given price.

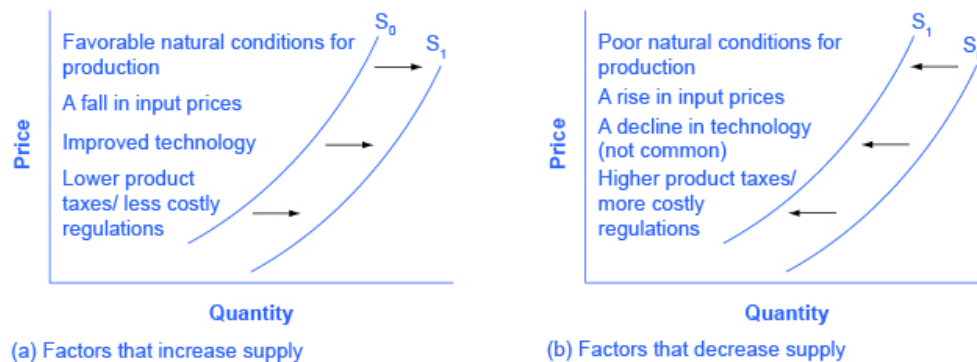
Government subsidies reduce the cost of production and increase supply at every given price, shifting supply to the right. The following Work It Out feature shows how this shift happens.

Summing Up Factors That Change Supply

Changes in the cost of inputs, natural disasters, new technologies, and the impact of government decisions all affect the cost of production. In turn, these factors affect how much firms are willing to supply at any given price.

[\[link\]](#) summarizes factors that change the supply of goods and services. Notice that a change in the price of the product itself is not among the factors that shift the supply curve. Although a change in price of a good or service typically causes a change in quantity supplied or a movement along the supply curve for that specific good or service, it does not cause the supply curve itself to shift.

Factors That Shift Supply Curves



- (a) A list of factors that can cause an increase in supply from S_0 to S_1 .
(b) The same factors, if their direction is reversed, can cause a decrease in supply from S_0 to S_1 .

Because demand and supply curves appear on a two-dimensional diagram with only price and quantity on the axes, an unwary visitor to the land of economics might be fooled into believing that economics is about only four

topics: demand, supply, price, and quantity. However, demand and supply are really “umbrella” concepts: demand covers all the factors that affect demand, and supply covers all the factors that affect supply. Factors other than price that affect demand and supply are included by using shifts in the demand or the supply curve. In this way, the two-dimensional demand and supply model becomes a powerful tool for analyzing a wide range of economic circumstances.

How do we tell?

When economists talk with one another, certain terms that have different meanings sound alike. In this section, you've uncovered an example of this.

When we talk about the Law of Demand, we are describing price as it relates to **quantity demanded**, without any other variables involved. When we allow these other variables (non-price factors such as changes in income, wealth, fashion, etc.) to enter the picture, we describe the creation of an entirely new demand curve, and we refer to that as a **shift in demand** or a **change in demand**.

So, when we say the **demand** will increase, what we're really saying is that there is a new demand curve that has been created as a result of a non-price factor being introduced. When we say the **quantity demanded** increased, what we're communicating is that we're talking about a movement along a demand curve: only price is influencing the new quantity demanded.

These same rules apply for a supply curve shift or a change in quantity supplied

Key Concepts and Summary

Economists often use the *ceteris paribus* or “other things being equal” assumption: while examining the economic impact of one event, all other factors remain unchanged for the purpose of the analysis. Factors that can shift the demand curve for goods and services, causing a different quantity to be demanded at any given price, include changes in tastes, population,

income, prices of substitute or complement goods, and expectations about future conditions and prices. Factors that can shift the supply curve for goods and services, causing a different quantity to be supplied at any given price, include input prices, natural conditions, changes in technology, and government taxes, regulations, or subsidies.

Self-Check Questions

Exercise:

Problem: Why do economists use the *ceteris paribus* assumption?

Solution:

To make it easier to analyze complex problems. *Ceteris paribus* allows you to look at the effect of one factor at a time on what it is you are trying to analyze. When you have analyzed all the factors individually, you add the results together to get the final answer.

Exercise:

Problem:

In an analysis of the market for paint, an economist discovers the facts listed below. State whether each of these changes will affect supply or demand, and in what direction.

- a. There have recently been some important cost-saving inventions in the technology for making paint.
- b. Paint is lasting longer, so that property owners need not repaint as often.
- c. Because of severe hailstorms, many people need to repaint now.
- d. The hailstorms damaged several factories that make paint, forcing them to close down for several months.

Solution:

- a. An improvement in technology that reduces the cost of production will cause an increase in supply. Alternatively, you can think of this as a reduction in price necessary for firms to supply any quantity. Either way, this can be shown as a rightward (or downward) shift in the supply curve.
- b. An improvement in product quality is treated as an increase in tastes or preferences, meaning consumers demand more paint at any price level, so demand increases or shifts to the right. If this seems counterintuitive, note that demand in the future for the longer-lasting paint will fall, since consumers are essentially shifting demand from the future to the present.
- c. An increase in need causes an increase in demand or a rightward shift in the demand curve.
- d. Factory damage means that firms are unable to supply as much in the present. Technically, this is an increase in the cost of production. Either way you look at it, the supply curve shifts to the left.

Exercise:

Problem:

Many changes are affecting the market for oil. Predict how each of the following events will affect the equilibrium price and quantity in the market for oil. In each case, state how the event will affect the supply and demand diagram. Create a sketch of the diagram if necessary.

- a. Cars are becoming more fuel efficient, and therefore get more miles to the gallon.
- b. The winter is exceptionally cold.
- c. A major discovery of new oil is made off the coast of Norway.
- d. The economies of some major oil-using nations, like Japan, slow down.
- e. A war in the Middle East disrupts oil-pumping schedules.
- f. Landlords install additional insulation in buildings.
- g. The price of solar energy falls dramatically.
- h. Chemical companies invent a new, popular kind of plastic made from oil.

Solution:

- a. More fuel-efficient cars means there is less need for gasoline. This causes a leftward shift in the demand for gasoline and thus oil. Since the demand curve is shifting down the supply curve, the equilibrium price and quantity both fall.
- b. Cold weather increases the need for heating oil. This causes a rightward shift in the demand for heating oil and thus oil. Since the demand curve is shifting up the supply curve, the equilibrium price and quantity both rise.
- c. A discovery of new oil will make oil more abundant. This can be shown as a rightward shift in the supply curve, which will cause a decrease in the equilibrium price along with an increase in the equilibrium quantity. (The supply curve shifts down the demand curve so price and quantity follow the law of demand. If price goes down, then the quantity goes up.)
- d. When an economy slows down, it produces less output and demands less input, including energy, which is used in the production of virtually everything. A decrease in demand for energy will be reflected as a decrease in the demand for oil, or a leftward shift in demand for oil. Since the demand curve is shifting down the supply curve, both the equilibrium price and quantity of oil will fall.
- e. Disruption of oil pumping will reduce the supply of oil. This leftward shift in the supply curve will show a movement up the demand curve, resulting in an increase in the equilibrium price of oil and a decrease in the equilibrium quantity.
- f. Increased insulation will decrease the demand for heating. This leftward shift in the demand for oil causes a movement down the supply curve, resulting in a decrease in the equilibrium price and quantity of oil.
- g. Solar energy is a substitute for oil-based energy. So if solar energy becomes cheaper, the demand for oil will decrease as consumers switch from oil to solar. The decrease in demand for oil will be shown as a leftward shift in the demand curve. As the

demand curve shifts down the supply curve, both equilibrium price and quantity for oil will fall.

- h. A new, popular kind of plastic will increase the demand for oil. The increase in demand will be shown as a rightward shift in demand, raising the equilibrium price and quantity of oil.

Review Questions

Exercise:

Problem:

When analyzing a market, how do economists deal with the problem that many factors that affect the market are changing at the same time?

Exercise:

Problem:

Name some factors that can cause a shift in the demand curve in markets for goods and services.

Exercise:

Problem:

Name some factors that can cause a shift in the supply curve in markets for goods and services.

Critical Thinking Questions

Exercise:

Problem:

Consider the demand for hamburgers. If the price of a substitute good (for example, hot dogs) increases and the price of a complement good (for example, hamburger buns) increases, can you tell for sure what will happen to the demand for hamburgers? Why or why not? Illustrate your answer with a graph.

Exercise:**Problem:**

How do you suppose the demographics of an aging population of “Baby Boomers” in the United States will affect the demand for milk? Justify your answer.

Exercise:**Problem:**

We know that a change in the price of a product causes a movement along the demand curve. Suppose consumers believe that prices will be rising in the future. How will that affect demand for the product in the present? Can you show this graphically?

Exercise:**Problem:**

Suppose there is soda tax to curb obesity. What should a reduction in the soda tax do to the supply of sodas and to the equilibrium price and quantity? Can you show this graphically? *Hint:* assume that the soda tax is collected from the sellers

Problems**Exercise:**

Problem:

[\[link\]](#) shows information on the demand and supply for bicycles, where the quantities of bicycles are measured in thousands.

Price	Qd	Qs
\$120	50	36
\$150	40	40
\$180	32	48
\$210	28	56
\$240	24	70

- What is the quantity demanded and the quantity supplied at a price of \$210?
- At what price is the quantity supplied equal to 48,000?
- Graph the demand and supply curve for bicycles. How can you determine the equilibrium price and quantity from the graph? How can you determine the equilibrium price and quantity from the table? What are the equilibrium price and equilibrium quantity?
- If the price was \$120, what would the quantities demanded and supplied be? Would a shortage or surplus exist? If so, how large would the shortage or surplus be?

Exercise:

Problem:

The computer market in recent years has seen many more computers sell at much lower prices. What shift in demand or supply is most likely to explain this outcome? Sketch a demand and supply diagram and explain your reasoning for each.

- a. A rise in demand
- b. A fall in demand
- c. A rise in supply
- d. A fall in supply

References

Landsburg, Steven E. *The Armchair Economist: Economics and Everyday Life*. New York: The Free Press. 2012. specifically Section IV: How Markets Work.

National Chicken Council. 2015. "Per Capita Consumption of Poultry and Livestock, 1965 to Estimated 2015, in Pounds." Accessed April 13, 2015. <http://www.nationalchickencouncil.org/about-the-industry/statistics/per-capita-consumption-of-poultry-and-livestock-1965-to-estimated-2012-in-pounds/>.

Wessel, David. "Saudi Arabia Fears \$40-a-Barrel Oil, Too." *The Wall Street Journal*. May 27, 2004, p. 42. <http://online.wsj.com/news/articles/SB108561000087822300>.

Glossary

ceteris paribus
other things being equal

complements

goods that are often used together so that consumption of one good tends to enhance consumption of the other

factors of production

the combination of labor, materials, and machinery that is used to produce goods and services; also called inputs

inferior good

a good in which the quantity demanded falls as income rises, and in which quantity demanded rises and income falls

inputs

the combination of labor, materials, and machinery that is used to produce goods and services; also called factors of production

normal good

a good in which the quantity demanded rises as income rises, and in which quantity demanded falls as income falls

shift in demand

when a change in some economic factor (other than price) causes a different quantity to be demanded at every price

shift in supply

when a change in some economic factor (other than price) causes a different quantity to be supplied at every price

substitute

a good that can replace another to some extent, so that greater consumption of one good can mean less of the other

Price Ceilings and Price Floors

By the end of this section, you will be able to:

- Explain price controls, price ceilings, and price floors
- Analyze demand and supply as a social adjustment mechanism

Controversy sometimes surrounds the prices and quantities established by demand and supply, especially for products that are considered necessities. In some cases, discontent over prices turns into public pressure on politicians, who may then pass legislation to prevent a certain price from climbing “too high” or falling “too low.”

The demand and supply model shows how people and firms will react to the incentives provided by these laws to control prices, in ways that will often lead to undesirable consequences. Alternative policy tools can often achieve the desired goals of price control laws, while avoiding at least some of their costs and tradeoffs.

Price Ceilings

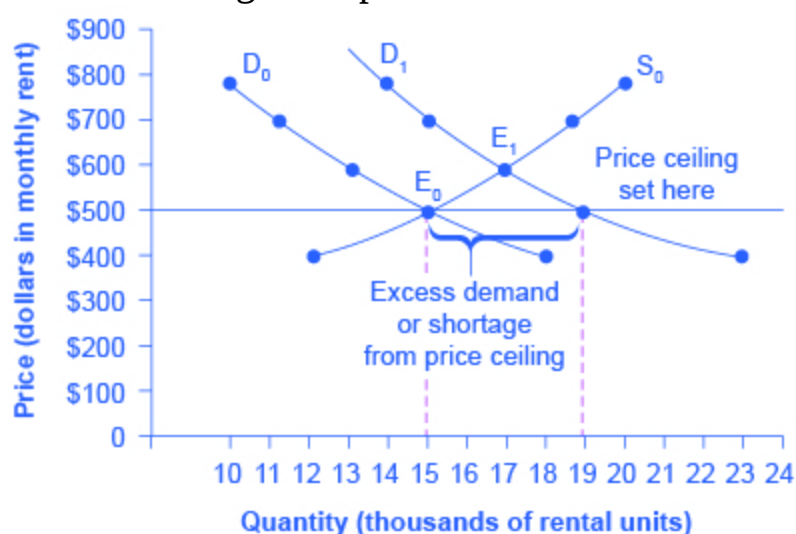
Laws that government enacts to regulate prices are called **Price controls**. Price controls come in two flavors. A **price ceiling** keeps a price from rising above a certain level (the “ceiling”), while a **price floor** keeps a price from falling below a certain level (the “floor”). This section uses the demand and supply framework to analyze price ceilings. The next section discusses price floors.

In many markets for goods and services, demanders outnumber suppliers. Consumers, who are also potential voters, sometimes unite behind a political proposal to hold down a certain price. In some cities, such as Albany, renters have pressed political leaders to pass rent control laws, a price ceiling that usually works by stating that rents can be raised by only a certain maximum percentage each year.

Rent control becomes a politically hot topic when rents begin to rise rapidly. Everyone needs an affordable place to live. Perhaps a change in tastes makes a certain suburb or town a more popular place to live. Perhaps

locally-based businesses expand, bringing higher incomes and more people into the area. Changes of this sort can cause a change in the demand for rental housing, as [\[link\]](#) illustrates. The original equilibrium (E_0) lies at the intersection of supply curve S_0 and demand curve D_0 , corresponding to an equilibrium price of \$500 and an equilibrium quantity of 15,000 units of rental housing. The effect of greater income or a change in tastes is to shift the demand curve for rental housing to the right, as shown by the data in [\[link\]](#) and the shift from D_0 to D_1 on the graph. In this market, at the new equilibrium E_1 , the price of a rental unit would rise to \$600 and the equilibrium quantity would increase to 17,000 units.

A Price Ceiling Example—Rent Control



The original intersection of demand and supply occurs at E_0 . If demand shifts from D_0 to D_1 , the new equilibrium would be at E_1 —unless a price ceiling prevents the price from rising. If the price is not permitted to rise, the quantity supplied remains at 15,000. However, after the change in demand, the quantity demanded rises to 19,000, resulting in a shortage.

Price	Original Quantity Supplied	Original Quantity Demanded	New Quantity Demanded
\$400	12,000	18,000	23,000
\$500	15,000	15,000	19,000
\$600	17,000	13,000	17,000
\$700	19,000	11,000	15,000
\$800	20,000	10,000	14,000

Rent Control

Suppose that a rent control law is passed to keep the price at the original equilibrium of \$500 for a typical apartment. In [\[link\]](#), the horizontal line at the price of \$500 shows the legally fixed maximum price set by the rent control law. However, the underlying forces that shifted the demand curve to the right are still there. At that price (\$500), the quantity supplied remains at the same 15,000 rental units, but the quantity demanded is 19,000 rental units. In other words, the quantity demanded exceeds the quantity supplied, so there is a shortage of rental housing. One of the ironies of price ceilings is that while the price ceiling was intended to help renters, there are actually fewer apartments rented out under the price ceiling (15,000 rental units) than would be the case at the market rent of \$600 (17,000 rental units).

Price ceilings do not simply benefit renters at the expense of landlords. Rather, some renters (or potential renters) lose their housing as landlords convert apartments to co-ops and condos. Even when the housing remains in the rental market, landlords tend to spend less on maintenance and on essentials like heating, cooling, hot water, and lighting. The first rule of economics is you do not get something for nothing—everything has an opportunity cost. So if renters get “cheaper” housing than the market requires, they tend to also end up with lower quality housing.

Price ceilings have been proposed for other products. For example, price ceilings to limit what producers can charge have been proposed in recent years for prescription drugs, doctor and hospital fees, the charges made by some automatic teller bank machines, and auto insurance rates. Price ceilings are enacted in an attempt to keep prices low for those who demand the product. But when the market price is not allowed to rise to the equilibrium level, quantity demanded exceeds quantity supplied, and thus a shortage occurs. Those who manage to purchase the product at the lower price given by the price ceiling will benefit, but sellers of the product will suffer, along with those who are not able to purchase the product at all. Quality is also likely to deteriorate.

Price Floors

A price floor is the lowest legal price that can be paid in markets for goods and services, labor, or financial capital. Perhaps the best-known example of a price floor is the minimum wage, which is based on the normative view that someone working full time ought to be able to afford a basic standard of living. The federal minimum wage at the end of 2014 was \$7.25 per hour, which yields an income for a single person slightly higher than the poverty line. As the cost of living rises over time, the Congress periodically raises the federal minimum wage.

Price floors are sometimes called “price supports,” because they support a price by preventing it from falling below a certain level. Around the world, many countries have passed laws to create agricultural price supports. Farm prices and thus farm incomes fluctuate, sometimes widely. So even if, on average, farm incomes are adequate, some years they can be quite low. The purpose of price supports is to prevent these swings.

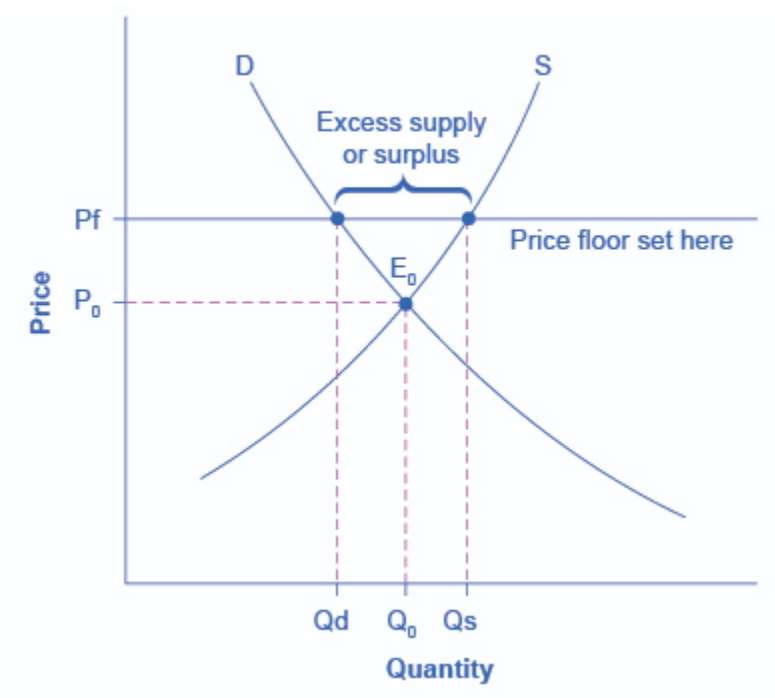
The most common way price supports work is that the government enters the market and buys up the product, adding to demand to keep prices higher than they otherwise would be. According to the Common Agricultural Policy reform passed in 2013, the European Union (EU) will spend about 60 billion euros per year, or 67 billion dollars per year, or roughly 38% of the EU budget, on price supports for Europe’s farmers from 2014 to 2020.

[\[link\]](#) illustrates the effects of a government program that assures a price above the equilibrium by focusing on the market for wheat in Europe. In the absence of government intervention, the price would adjust so that the quantity supplied would equal the quantity demanded at the equilibrium point E_0 , with price P_0 and quantity Q_0 . However, policies to keep prices high for farmers keeps the price above what would have been the market equilibrium level—the price P_f shown by the dashed horizontal line in the diagram. The result is a quantity supplied in excess of the quantity demanded (Q_d). When quantity supplied exceeds quantity demanded, a surplus exists.

The high-income areas of the world, including the United States, Europe, and Japan, are estimated to spend roughly \$1 billion per day in supporting their farmers. If the government is willing to purchase the excess supply (or to provide payments for others to purchase it), then farmers will benefit from the price floor, but taxpayers and consumers of food will pay the costs. Numerous proposals have been offered for reducing farm subsidies. In many countries, however, political support for subsidies for farmers remains strong. Either because this is viewed by the population as supporting the traditional rural way of life or because of the lobbying power of the agribusiness industry.

For more detail on the effects price ceilings and floors have on demand and supply, see the following Clear It Up feature.

European Wheat Prices: A Price Floor Example



The intersection of demand (D) and supply (S) would be at the equilibrium point E_0 .

However, a price floor set at P_f holds the price above E_0 and prevents it from falling.

The result of the price floor is that the quantity supplied Q_s exceeds the quantity demanded Q_d . There is excess supply, also called a surplus.

Note:

Do price ceilings and floors change demand or supply?

Neither price ceilings nor price floors cause demand or supply to change. They simply set a price that limits what can be legally charged in the market. Remember, changes in price do not cause demand or supply to change. Price ceilings and price floors can cause a different choice of quantity demanded along a demand curve, but they do not move the

demand curve. Price controls can cause a different choice of quantity supplied along a supply curve, but they do not shift the supply curve.

Key Concepts and Summary

Price ceilings prevent a price from rising above a certain level. When a price ceiling is set below the equilibrium price, quantity demanded will exceed quantity supplied, and excess demand or shortages will result. Price floors prevent a price from falling below a certain level. When a price floor is set above the equilibrium price, quantity supplied will exceed quantity demanded, and excess supply or surpluses will result. Price floors and price ceilings often lead to unintended consequences.

Self-Check Questions

Exercise:

Problem:

What is the effect of a price ceiling on the quantity demanded of the product? What is the effect of a price ceiling on the quantity supplied? Why exactly does a price ceiling cause a shortage?

Solution:

A price ceiling (which is below the equilibrium price) will cause the quantity demanded to rise and the quantity supplied to fall. This is why a price ceiling creates a shortage.

Exercise:

Problem: Does a price ceiling change the equilibrium price?

Solution:

A price ceiling is just a legal restriction. Equilibrium is an economic condition. People may or may not obey the price ceiling, so the actual price may be at or above the price ceiling, but the price ceiling does not change the equilibrium price.

Exercise:

Problem:

What would be the impact of imposing a price floor below the equilibrium price?

Solution:

A price ceiling is a legal maximum price, but a price floor is a legal minimum price and, consequently, it would leave room for the price to rise to its equilibrium level. In other words, a price floor below equilibrium will not be binding and will have no effect.

Review Questions

Exercise:

Problem:

Does a price ceiling attempt to make a price higher or lower?

Exercise:

Problem:

How does a price ceiling set below the equilibrium level affect quantity demanded and quantity supplied?

Exercise:

Problem: Does a price floor attempt to make a price higher or lower?

Exercise:

Problem:

How does a price floor set above the equilibrium level affect quantity demanded and quantity supplied?

Critical Thinking Questions**Exercise:****Problem:**

Most government policy decisions have winners and losers. What are the effects of raising the minimum wage? It is more complex than simply producers lose and workers gain. Who are the winners and who are the losers, and what exactly do they win and lose? To what extent does the policy change achieve its goals?

Exercise:**Problem:**

Agricultural price supports result in governments holding large inventories of agricultural products. Why do you think the government cannot simply give the products away to poor people?

Exercise:**Problem:**

Can you propose a policy that would induce the market to supply more rental housing units?

Problems**Exercise:**

Problem:

A low-income country decides to set a price ceiling on bread so it can make sure that bread is affordable to the poor. The conditions of demand and supply are given in [\[link\]](#). What are the equilibrium price and equilibrium quantity before the price ceiling? What will the excess demand or the shortage (that is, quantity demanded minus quantity supplied) be if the price ceiling is set at \$2.40? At \$2.00? At \$3.60?

Price	Qd	Qs
\$1.60	9,000	5,000
\$2.00	8,500	5,500
\$2.40	8,000	6,400
\$2.80	7,500	7,500
\$3.20	7,000	9,000
\$3.60	6,500	11,000
\$4.00	6,000	15,000

Glossary

price ceiling

a legal maximum price

price control

government laws to regulate prices instead of letting market forces determine prices

price floor

a legal minimum price

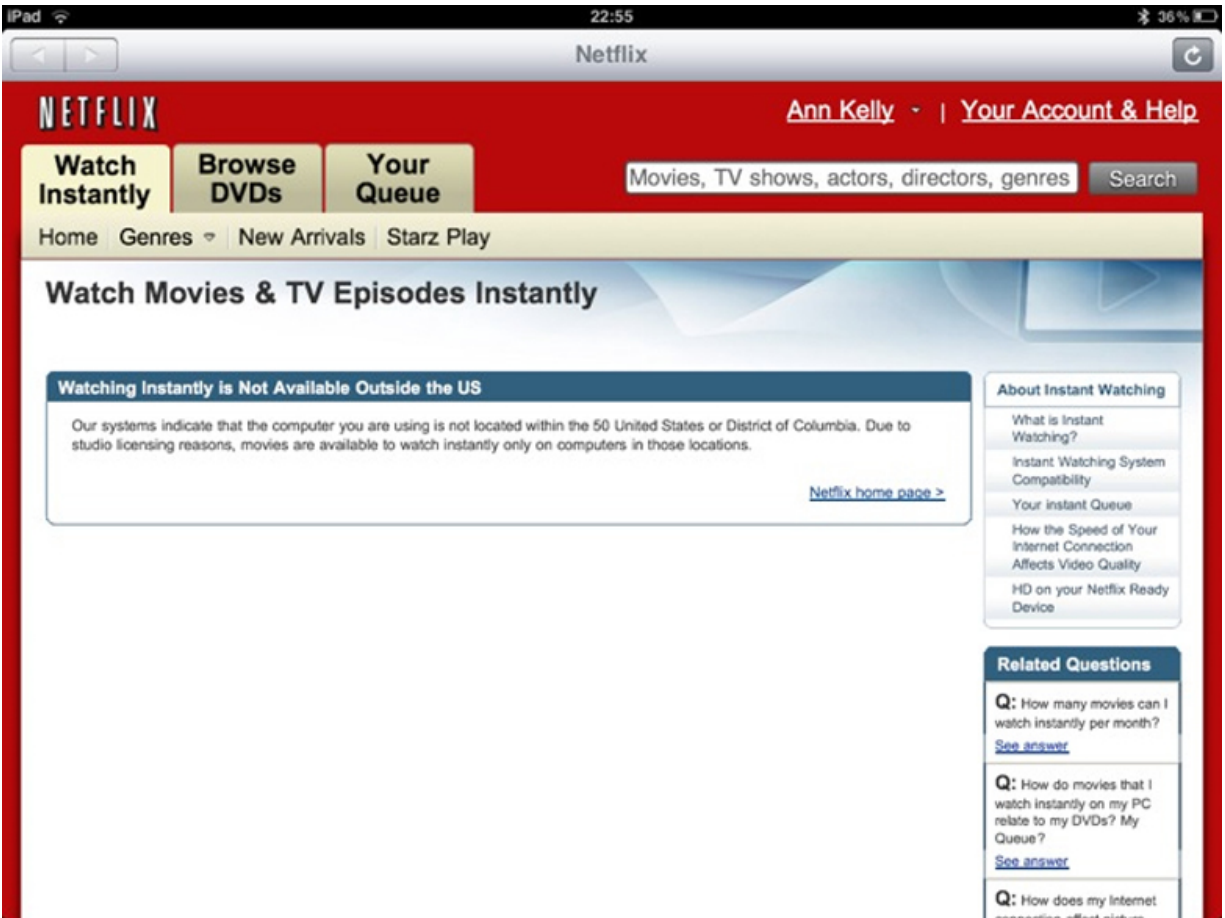
total surplus

see social surplus

Introduction to Elasticity
class="introduction"
Netflix On-Demand Media

Netflix, Inc. is
an American
provider of
on-demand
Internet
streaming
media to
many
countries
around the
world,
including the
United States,
and of flat
rate DVD-by-
mail in the
United States.

(Credit:
modification
of work by
Traci
Lawson/Flick
r Creative
Commons)



Note:

That Will Be How Much?

Imagine going to your favorite coffee shop and having the waiter inform you the pricing has changed. Instead of \$3 for a cup of coffee, you will now be charged \$2 for coffee, \$1 for creamer, and \$1 for your choice of sweetener. If you pay your usual \$3 for a cup of coffee, you must choose between creamer and sweetener. If you want both, you now face an extra charge of \$1. Sound absurd? Well, that is the situation Netflix customers found themselves in—a 60% price hike to retain the same service in 2011. In early 2011, Netflix consumers paid about \$10 a month for a package consisting of streaming video and DVD rentals. In July 2011, the company announced a packaging change. Customers wishing to retain both streaming video and DVD rental would be charged \$15.98 per month, a price increase of about 60%. In 2014, Netflix also raised its streaming

video subscription price from \$7.99 to \$8.99 per month for new U.S. customers. The company also changed its policy of 4K streaming content from \$9.00 to \$12.00 per month that year.

How would customers of the 18-year-old firm react? Would they abandon Netflix? Would the ease of access to other venues make a difference in how consumers responded to the Netflix price change? The answers to those questions will be explored in this chapter: the change in quantity with respect to a change in price, a concept economists call elasticity.

Note:

Introduction to Elasticity

In this chapter, you will learn about:

- Price Elasticity of Demand and Price Elasticity of Supply
- Polar Cases of Elasticity and Constant Elasticity
- Elasticity and Pricing
- Elasticity in Areas Other Than Price

Anyone who has studied economics knows the law of demand: a higher price will lead to a lower quantity demanded. What you may not know is how much lower the quantity demanded will be. Similarly, the law of supply shows that a higher price will lead to a higher quantity supplied. The question is: How much higher? This chapter will explain how to answer these questions and why they are critically important in the real world.

To find answers to these questions, we need to understand the concept of elasticity. **Elasticity** is an economics concept that measures responsiveness of one variable to changes in another variable. Suppose you drop two items from a second-floor balcony. The first item is a tennis ball. The second item is a brick. Which will bounce higher? Obviously, the tennis ball. We would say that the tennis ball has greater elasticity.

Consider an economic example. Cigarette taxes are an example of a “sin tax,” a tax on something that is bad for you, like alcohol. Cigarettes are taxed at the state and national levels. State taxes range from a low of 17 cents per pack in Missouri to \$4.35 per pack in New York. The average state cigarette tax is \$1.51 per pack. The 2014 federal tax rate on cigarettes was \$1.01 per pack, but in 2015 the Obama Administration proposed raising the federal tax nearly a dollar to \$1.95 per pack. The key question is: How much would cigarette purchases decline?

Taxes on cigarettes serve two purposes: to raise tax revenue for government and to discourage consumption of cigarettes. However, if a higher cigarette tax discourages consumption by quite a lot, meaning a greatly reduced quantity of cigarettes is sold, then the cigarette tax on each pack will not raise much revenue for the government. Alternatively, a higher cigarette tax that does not discourage consumption by much will actually raise more tax revenue for the government. Thus, when a government agency tries to calculate the effects of altering its cigarette tax, it must analyze how much the tax affects the quantity of cigarettes consumed. This issue reaches beyond governments and taxes; every firm faces a similar issue. Every time a firm considers raising the price that it charges, it must consider how much a price increase will reduce the quantity demanded of what it sells. Conversely, when a firm puts its products on sale, it must expect (or hope) that the lower price will lead to a significantly higher quantity demanded.

Price Elasticity of Demand and Price Elasticity of Supply

By the end of this section, you will be able to:

- Calculate the price elasticity of demand
- Calculate the price elasticity of supply

Both the demand and supply curve show the relationship between price and the number of units demanded or supplied. **Price elasticity** is the ratio between the percentage change in the quantity demanded (Qd) or supplied (Qs) and the corresponding percent change in price. The **price elasticity of demand** is the percentage change in the quantity *demanded* of a good or service divided by the percentage change in the price. The **price elasticity of supply** is the percentage change in quantity *supplied* divided by the percentage change in price.

Elasticities can be usefully divided into three broad categories: elastic, inelastic, and unitary. An **elastic demand** or **elastic supply** is one in which the elasticity is greater than one, indicating a high responsiveness to changes in price. Elasticities that are less than one indicate low responsiveness to price changes and correspond to **inelastic demand** or **inelastic supply**. **Unitary elasticities** indicate proportional responsiveness of either demand or supply, as summarized in [\[link\]](#).

If . . .	Then . . .	And It Is Called . . .
% change in quantity > % change in price	$\frac{\% \text{ change in quantity}}{\% \text{ change in price}} > 1$	Elastic
% change in quantity = % change in price	$\frac{\% \text{ change in quantity}}{\% \text{ change in price}} = 1$	Unitary
% change in quantity < % change in price	$\frac{\% \text{ change in quantity}}{\% \text{ change in price}} < 1$	Inelastic

Elastic, Inelastic, and Unitary: Three Cases of Elasticity

Key Concepts and Summary

Price elasticity measures the responsiveness of the quantity demanded or supplied of a good to a change in its price. It is computed as the percentage change in quantity demanded (or supplied) divided by the percentage change in price. Elasticity can be described as elastic (or very responsive), unit elastic, or inelastic (not very responsive). Elastic demand or supply

curves indicate that quantity demanded or supplied respond to price changes in a greater than proportional manner. An inelastic demand or supply curve is one where a given percentage change in price will cause a smaller percentage change in quantity demanded or supplied. A unitary elasticity means that a given percentage change in price leads to an equal percentage change in quantity demanded or supplied.

Self-Check Questions

Exercise:

Problem:

From the data shown in [\[link\]](#) about demand for smart phones, calculate the price elasticity of demand from: point B to point C, point D to point E, and point G to point H. Classify the elasticity at each point as elastic, inelastic, or unit elastic.

Points	P	Q
A	60	3,000
B	70	2,800
C	80	2,600
D	90	2,400
E	100	2,200
F	110	2,000
G	120	1,800
H	130	1,600

Solution:

From point B to point C, price rises from \$70 to \$80, and Qd decreases from 2,800 to 2,600. So:

Equation:

$$\begin{aligned}
 \% \text{ change in quantity} &= \frac{2600-2800}{(2600+2800) \div 2} \times 100 \\
 &= \frac{-200}{2700} \times 100 \\
 &= -7.41 \\
 \% \text{ change in price} &= \frac{80-70}{(80+70) \div 2} \times 100 \\
 &= \frac{10}{75} \times 100 \\
 &= 13.33 \\
 \text{Elasticity of Demand} &= \frac{-7.41\%}{13.33\%} \\
 &= 0.56
 \end{aligned}$$

The demand curve is inelastic in this area; that is, its elasticity value is less than one.

Answer from Point D to point E:

Equation:

$$\begin{aligned}
 \% \text{ change in quantity} &= \frac{2200-2400}{(2200+2400) \div 2} \times 100 \\
 &= \frac{-200}{2300} \times 100 \\
 &= -8.7 \\
 \% \text{ change in price} &= \frac{100-90}{(100+90) \div 2} \times 100 \\
 &= \frac{10}{95} \times 100 \\
 &= 10.53 \\
 \text{Elasticity of Demand} &= \frac{-8.7\%}{10.53\%} \\
 &= 0.83
 \end{aligned}$$

The demand curve is inelastic in this area; that is, its elasticity value is less than one.

Answer from Point G to point H:

Equation:

$$\begin{aligned}
 \% \text{ change in quantity} &= \frac{1600-1800}{1700} \times 100 \\
 &= \frac{-200}{1700} \times 100 \\
 &= -11.76 \\
 \% \text{ change in price} &= \frac{130-120}{125} \times 100 \\
 &= \frac{10}{125} \times 100 \\
 &= 8.00 \\
 \text{Elasticity of Demand} &= \frac{-11.76\%}{8.00\%} \\
 &= -1.47
 \end{aligned}$$

The demand curve is elastic in this interval.

Exercise:

Problem:

From the data shown in [\[link\]](#) about supply of alarm clocks, calculate the price elasticity of supply from: point J to point K, point L to point M, and point N to point P. Classify the elasticity at each point as elastic, inelastic, or unit elastic.

Point	Price	Quantity Supplied
J	\$8	50
K	\$9	70
L	\$10	80
M	\$11	88
N	\$12	95
P	\$13	100

Solution:

From point J to point K, price rises from \$8 to \$9, and quantity rises from 50 to 70. So:

Equation:

$$\begin{aligned}\% \text{ change in quantity} &= \frac{70-50}{(70+50) \div 2} \times 100 \\ &= \frac{20}{60} \times 100 \\ &= 33.33\end{aligned}$$

$$\begin{aligned}\% \text{ change in price} &= \frac{\$9-\$8}{(\$9+\$8) \div 2} \times 100 \\ &= \frac{1}{\text{mtd}} \times 100 \\ &= 11.76\end{aligned}$$

$$\begin{aligned}\text{Elasticity of Supply} &= \frac{33.33\%}{11.76\%} \\ &= 2.83\end{aligned}$$

The supply curve is elastic in this area; that is, its elasticity value is greater than one.

From point L to point M, the price rises from \$10 to \$11, while the Qs rises from 80 to 88:

Equation:

$$\begin{aligned}\% \text{ change in quantity} &= \frac{88-80}{(88+80) \div 2} \times 100 \\ &= \frac{8}{84} \times 100 \\ &= 9.52\end{aligned}$$

$$\begin{aligned}\% \text{ change in price} &= \frac{\$11-\$10}{(\$11+\$10) \div 2} \times 100 \\ &= \frac{1}{10.5} \times 100 \\ &= 9.52\end{aligned}$$

$$\begin{aligned}\text{Elasticity of Demand} &= \frac{9.52\%}{9.52\%} \\ &= 1.0\end{aligned}$$

The supply curve has unitary elasticity in this area.

From point N to point P, the price rises from \$12 to \$13, and Qs rises from 95 to 100:

Equation:

$$\begin{aligned}\% \text{ change in quantity} &= \frac{100-95}{(100+95) \div 2} \times 100 \\ &= \frac{5}{97.5} \times 100 \\ &= 5.13\end{aligned}$$

$$\begin{aligned}\% \text{ change in price} &= \frac{\$13-\$12}{(\$13+\$12) \div 2} \times 100 \\ &= \frac{1}{12.5} \times 100 \\ &= 8.0\end{aligned}$$

$$\begin{aligned}\text{Elasticity of Supply} &= \frac{5.13\%}{8.0\%} \\ &= 0.64\end{aligned}$$

The supply curve is inelastic in this region of the supply curve.

Review Questions

Exercise:

Problem: What is the formula for calculating elasticity?

Exercise:

Problem:

What is the price elasticity of demand? Can you explain it in your own words?

Exercise:

Problem: What is the price elasticity of supply? Can you explain it in your own words?

Critical Thinking Questions

Exercise:

Problem:

Transatlantic air travel in business class has an estimated elasticity of demand of 0.40 less than transatlantic air travel in economy class, with an estimated price elasticity of 0.62. Why do you think this is the case?

Exercise:

Problem:

What is the relationship between price elasticity and position on the demand curve? For example, as you move up the demand curve to higher prices and lower quantities, what happens to the measured elasticity? How would you explain that?

Problems**Exercise:****Problem:**

The equation for a demand curve is $P = 48 - 3Q$. What is the elasticity in moving from a quantity of 5 to a quantity of 6?

Exercise:**Problem:**

The equation for a demand curve is $P = 2/Q$. What is the elasticity of demand as price falls from 5 to 4? What is the elasticity of demand as the price falls from 9 to 8? Would you expect these answers to be the same?

Exercise:**Problem:**

The equation for a supply curve is $4P = Q$. What is the elasticity of supply as price rises from 3 to 4? What is the elasticity of supply as the price rises from 7 to 8? Would you expect these answers to be the same?

Exercise:**Problem:**

The equation for a supply curve is $P = 3Q - 8$. What is the elasticity in moving from a price of 4 to a price of 7?

Glossary**elastic demand**

when the elasticity of demand is greater than one, indicating a high responsiveness of quantity demanded or supplied to changes in price

elastic supply

when the elasticity of either supply is greater than one, indicating a high responsiveness of quantity demanded or supplied to changes in price

elasticity

an economics concept that measures responsiveness of one variable to changes in another variable

inelastic demand

when the elasticity of demand is less than one, indicating that a 1 percent increase in price paid by the consumer leads to less than a 1 percent change in purchases (and vice versa); this indicates a low responsiveness by consumers to price changes

inelastic supply

when the elasticity of supply is less than one, indicating that a 1 percent increase in price paid to the firm will result in a less than 1 percent increase in production by the firm; this indicates a low responsiveness of the firm to price increases (and vice versa if prices drop)

price elasticity

the relationship between the percent change in price resulting in a corresponding percentage change in the quantity demanded or supplied

price elasticity of demand

percentage change in the quantity *demanded* of a good or service divided the percentage change in price

price elasticity of supply

percentage change in the quantity *supplied* divided by the percentage change in price

unitary elasticity

when the calculated elasticity is equal to one indicating that a change in the price of the good or service results in a proportional change in the quantity demanded or supplied

Elasticity and Pricing

By the end of this section, you will be able to:

- Analyze how price elasticities impact revenue
- Evaluate how elasticity can cause shifts in demand and supply
- Predict how the long-run and short-run impacts of elasticity affect equilibrium
- Explain how the elasticity of demand and supply determine the incidence of a tax on buyers and sellers

Studying elasticities is useful for a number of reasons, pricing being most important. Let's explore how elasticity relates to revenue and pricing, both in the long run and short run. But first, let's look at the elasticities of some common goods and services.

[\[link\]](#) shows a selection of demand elasticities for different goods and services drawn from a variety of different studies by economists, listed in order of increasing elasticity.

Goods and Services	Elasticity of Price
Housing	0.12
Transatlantic air travel (economy class)	0.12
Rail transit (rush hour)	0.15
Electricity	0.20
Taxi cabs	0.22

Goods and Services	Elasticity of Price
Gasoline	0.35
Transatlantic air travel (first class)	0.40
Wine	0.55
Beef	0.59
Transatlantic air travel (business class)	0.62
Kitchen and household appliances	0.63
Cable TV (basic rural)	0.69
Chicken	0.64
Soft drinks	0.70
Beer	0.80
New vehicle	0.87
Rail transit (off-peak)	1.00
Computer	1.44
Cable TV (basic urban)	1.51
Cable TV (premium)	1.77
Restaurant meals	2.27

Some Selected Elasticities of Demand

Note that necessities such as housing and electricity are inelastic, while items that are not necessities such as restaurant meals are more price-sensitive. If the price of the restaurant meal increases by 10%, the quantity demanded will decrease by 22.7%. A 10% increase in the price of housing will cause a slight decrease of 1.2% in the quantity of housing demanded.

Does Raising Price Bring in More Revenue?

Imagine that a band on tour is playing in an indoor arena with 15,000 seats. To keep this example simple, assume that the band keeps all the money from ticket sales. Assume further that the band pays the costs for its appearance, but that these costs, like travel, setting up the stage, and so on, are the same regardless of how many people are in the audience. Finally, assume that all the tickets have the same price. (The same insights apply if ticket prices are more expensive for some seats than for others, but the calculations become more complicated.) The band knows that it faces a downward-sloping demand curve; that is, if the band raises the price of tickets, it will sell fewer tickets. How should the band set the price for tickets to bring in the most total revenue, which in this example, because costs are fixed, will also mean the highest profits for the band? Should the band sell more tickets at a lower price or fewer tickets at a higher price?

The key concept in thinking about collecting the most revenue is the price elasticity of demand. Total revenue is price times the quantity of tickets sold. Imagine that the band starts off thinking about a certain price, which will result in the sale of a certain quantity of tickets. The three possibilities are laid out in [\[link\]](#). If demand is elastic at that price level, then the band should cut the price, because the percentage drop in price will result in an even larger percentage increase in the quantity sold—thus raising total revenue. However, if demand is inelastic at that original quantity level, then the band should raise the price of tickets, because a certain percentage increase in price will result in a smaller percentage decrease in the quantity sold—and total revenue will rise. If demand has a unitary elasticity at that quantity, then a moderate percentage change in the price will be offset by an equal percentage change in quantity—so the band will earn the same revenue whether it (moderately) increases or decreases the price of tickets.

If Demand Is ...	Then ...	Therefore . . .
Elastic	$\% \text{ change in } Q_d > \% \text{ change in } P$	A given % rise in P will be more than offset by a larger % fall in Q so that total revenue ($P \times Q$) falls.
Unitary	$\% \text{ change in } Q_d = \% \text{ change in } P$	A given % rise in P will be exactly offset by an equal % fall in Q so that total revenue ($P \times Q$) is unchanged.
Inelastic	$\% \text{ change in } Q_d < \% \text{ change in } P$	A given % rise in P will cause a smaller % fall in Q so that total revenue ($P \times Q$) rises.

Will the Band Earn More Revenue by Changing Ticket Prices?

What if the band keeps cutting price, because demand is elastic, until it reaches a level where all 15,000 seats in the available arena are sold? If

demand remains elastic at that quantity, the band might try to move to a bigger arena, so that it could cut ticket prices further and see a larger percentage increase in the quantity of tickets sold. Of course, if the 15,000-seat arena is all that is available or if a larger arena would add substantially to costs, then this option may not work.

Conversely, a few bands are so famous, or have such fanatical followings, that demand for tickets may be inelastic right up to the point where the arena is full. These bands can, if they wish, keep raising the price of tickets. Ironically, some of the most popular bands could make more revenue by setting prices so high that the arena is not filled—but those who buy the tickets would have to pay very high prices. However, bands sometimes choose to sell tickets for less than the absolute maximum they might be able to charge, often in the hope that fans will feel happier and spend more on recordings, T-shirts, and other paraphernalia.

Key Concepts and Summary

In the market for goods and services, the elasticity of a good or service tends to be greater (more responsive) for those with close substitutes, luxuries, and those things that have a small price compared to income or wealth. In addition, when we have more time to make a decision, we tend to respond in a more price sensitive manner in part, because we have time to find alternatives or to shop around. In addition, categories of goods or services tend to be more price inelastic than individual items within the category (our demand for apples tends to be more price elastic than our demand for the category of food).

Knowing the price elasticity of demand for a good or service helps us decide whether an increase in price will raise our total revenue, or lower it. (remember, total revenue = price x quantity)

Self-Check Questions

Exercise:

Problem:

The federal government decides to require that automobile manufacturers install new anti-pollution equipment that costs \$2,000 per car. Under what conditions can carmakers pass almost all of this cost along to car buyers? Under what conditions can carmakers pass very little of this cost along to car buyers?

Solution:

Carmakers can pass this cost along to consumers if the demand for these cars is inelastic. If the demand for these cars is elastic, then the manufacturer must pay for the equipment.

Exercise:**Problem:**

Suppose you are in charge of sales at a pharmaceutical company, and your firm has a new drug that causes bald men to grow hair. Assume that the company wants to earn as much revenue as possible from this drug. If the elasticity of demand for your company's product at the current price is 1.4, would you advise the company to raise the price, lower the price, or to keep the price the same? What if the elasticity were 0.6? What if it were 1? Explain your answer.

Solution:

If the elasticity is 1.4 at current prices, you would advise the company to lower its price on the product, since a decrease in price will be offset by the increase in the amount of the drug sold. If the elasticity were 0.6, then you would advise the company to increase its price. Increases in price will offset the decrease in number of units sold, but increase your total revenue. If elasticity is 1, the total revenue is already maximized, and you would advise that the company maintain its current price level.

Review Questions

Exercise:

Problem:

If demand is elastic, will shifts in supply have a larger effect on equilibrium quantity or on price?

Exercise:

Problem:

If demand is inelastic, will shifts in supply have a larger effect on equilibrium price or on quantity?

Exercise:

Problem:

If supply is elastic, will shifts in demand have a larger effect on equilibrium quantity or on price?

Exercise:

Problem:

If supply is inelastic, will shifts in demand have a larger effect on equilibrium price or on quantity?

Exercise:

Problem:

Would you usually expect elasticity of demand or supply to be higher in the short run or in the long run? Why?

Exercise:

Problem:

Under which circumstances does the tax burden fall entirely on consumers?

Critical Thinking Questions

Exercise:

Problem:

Would you expect supply to play a more significant role in determining the price of a basic necessity like food or a luxury like perfume? Explain. *Hint:* Think about how the price elasticity of demand will differ between necessities and luxuries.

Exercise:

Problem:

A city has built a bridge over a river and it decides to charge a toll to everyone who crosses. For one year, the city charges a variety of different tolls and records information on how many drivers cross the bridge. The city thus gathers information about elasticity of demand. If the city wishes to raise as much revenue as possible from the tolls, where will the city decide to charge a toll: in the inelastic portion of the demand curve, the elastic portion of the demand curve, or the unit elastic portion? Explain.

Exercise:

Problem:

In a market where the supply curve is perfectly inelastic, how does an excise tax affect the price paid by consumers and the quantity bought and sold?

Problems

Exercise:

Problem:

Assume that the supply of low-skilled workers is fairly elastic, but the employers' demand for such workers is fairly inelastic. If the policy goal is to expand employment for low-skilled workers, is it better to focus on policy tools to shift the supply of unskilled labor or on tools to shift the demand for unskilled labor? What if the policy goal is to raise wages for this group? Explain your answers with supply and demand diagrams.

Glossary

tax incidence

manner in which the tax burden is divided between buyers and sellers

Introduction to the Macroeconomic Perspective

class="introduction"

The Great Depression

At times, such as when many people are in need of government assistance, it is easy to tell how the economy is doing.

This photograph shows people lined up during the Great Depression, waiting for relief checks. At

other times, when some are doing well and others are not, it is more difficult to ascertain how the economy of a country is doing.

(Credit: modification of work by the U.S.

Library of Congress/Wikimedia Commons)

**Note:****How is the Economy Doing? How Does One Tell?**

The 1990s were boom years for the U.S. economy. The late 2000s, from 2007 to 2014 were not. What causes the economy to expand or contract? Why do businesses fail when they are making all the right decisions? Why do workers lose their jobs when they are hardworking and productive? Are bad economic times a failure of the market system? Are they a failure of the government? These are all questions of macroeconomics, which we will begin to address in this chapter. We will not be able to answer all of these questions here, but we will start with the basics: How is the economy doing? How can we tell?

The macro economy includes all buying and selling, all production and consumption; everything that goes on in every market in the economy. How can we get a handle on that? The answer begins more than 80 years ago, during the Great Depression. President Franklin D. Roosevelt and his economic advisers knew things were bad—but how could they express and measure just how bad it was? An economist named Simon Kuznets, who later won the Nobel Prize for his work, came up with a way to track what the entire economy is producing. The result—gross domestic product

(GDP)—remains our basic measure of macroeconomic activity. In this chapter, you will learn how GDP is constructed, how it is used, and why it is so important.

Note:

Introduction to the Macroeconomic Perspective

In this chapter, you will learn about:

- Measuring the Size of the Economy: Gross Domestic Product
- Adjusting Nominal Values to Real Values
- Tracking Real GDP over Time
- Comparing GDP among Countries
- How Well GDP Measures the Well-Being of Society

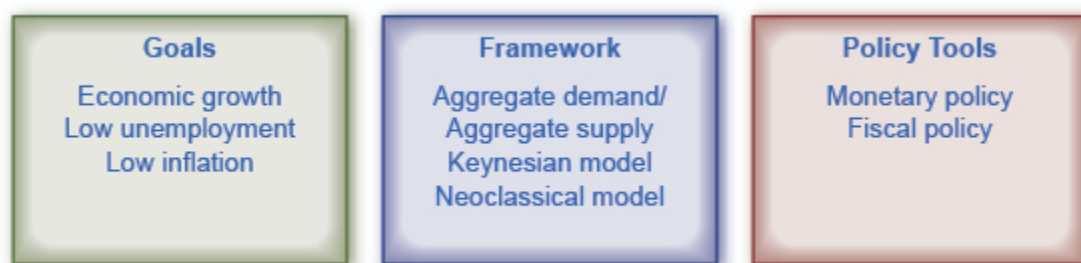
Macroeconomics focuses on the economy as a whole (or on whole economies as they interact). What causes recessions? What makes unemployment stay high when recessions are supposed to be over? Why do some countries grow faster than others? Why do some countries have higher standards of living than others? These are all questions that macroeconomics addresses. Macroeconomics involves adding up the economic activity of all households and all businesses in all markets to get the overall demand and supply in the economy. However, when we do that, something curious happens. It is not unusual that what results at the macro level is different from the sum of the microeconomic parts. Indeed, what seems sensible from a microeconomic point of view can have unexpected or counterproductive results at the macroeconomic level. Imagine that you are sitting at an event with a large audience, like a live concert or a basketball game. A few people decide that they want a better view, and so they stand up. However, when these people stand up, they block the view for other people, and the others need to stand up as well if they wish to see. Eventually, nearly everyone is standing up, and as a result, no one can see much better than before. The rational decision of some individuals at the

micro level—to stand up for a better view—ended up being self-defeating at the macro level. This is not macroeconomics, but it is an apt analogy.

Macroeconomics is a rather massive subject. How are we going to tackle it? [\[link\]](#) illustrates the structure we will use. We will study macroeconomics from three different perspectives:

1. What are the macroeconomic goals? (Macroeconomics as a discipline does not have goals, but we do have goals for the macro economy.)
2. What are the frameworks economists can use to analyze the macroeconomy?
3. Finally, what are the policy tools governments can use to manage the macroeconomy?

Macroeconomic Goals, Framework, and Policies



This chart shows what macroeconomics is about. The box on the left indicates a consensus of what are the most important goals for the macro economy, the middle box lists the frameworks economists use to analyze macroeconomic changes (such as inflation or recession), and the box on the right indicates the two tools the federal government uses to influence the macro economy.

Goals

In thinking about the overall health of the macroeconomy, it is useful to consider three primary goals: economic growth, low unemployment, and low inflation.

- Economic growth ultimately determines the prevailing standard of living in a country. Economic growth is measured by the percentage change in real (inflation-adjusted) gross domestic product. A growth rate of more than 3% is considered good.
- Unemployment, as measured by the unemployment rate, is the percentage of people in the labor force who do not have a job. When people lack jobs, the economy is wasting a precious resource—labor, and the result is lower goods and services produced. Unemployment, however, is more than a statistic—it represents people’s livelihoods. While measured unemployment is unlikely to ever be zero, a measured unemployment rate of 5% or less is considered low (good).
- Inflation is a sustained increase in the overall level of prices, and is measured by the consumer price index. If many people face a situation where the prices that they pay for food, shelter, and healthcare are rising much faster than the wages they receive for their labor, there will be widespread unhappiness as their standard of living declines. For that reason, low inflation—an inflation rate of 1–2%—is a major goal.

Frameworks

As you learn in the micro part of this book, principal tools used by economists are theories and models (see [Welcome to Economics!](#) for more on this). In microeconomics, we used the theories of supply and demand; in macroeconomics, we use the theories of aggregate demand (AD) and aggregate supply (AS). This book presents two perspectives on macroeconomics: the Neoclassical perspective and the Keynesian perspective, each of which has its own version of AD and AS. Between the two perspectives, you will obtain a good understanding of what drives the macroeconomy.

Policy Tools

National governments have two tools for influencing the macroeconomy. The first is monetary policy, which involves managing the money supply

and interest rates. The second is fiscal policy, which involves changes in government spending/purchases and taxes.

Each of the items in [\[link\]](#) will be explained in detail in one or more other chapters. As you learn these things, you will discover that the goals and the policy tools are in the news almost every day.

Measuring the Size of the Economy: Gross Domestic Product

By the end of this section, you will be able to:

- Identify the components of GDP on the demand side and on the supply side
- Evaluate how gross domestic product (GDP) is measured
- Contrast and calculate GDP, net exports, and net national product

Macroeconomics is an empirical subject, so the first step toward understanding it is to measure the economy.

How large is the U.S. economy? The size of a nation's overall economy is typically measured by its **gross domestic product (GDP)**, which is the value of all final goods and services produced within a country in a given year. The measurement of GDP involves counting up the production of millions of different goods and services—smart phones, cars, music downloads, computers, steel, bananas, college educations, and all other new goods and services produced in the current year—and summing them into a total dollar value. This task is straightforward: take the quantity of everything produced, multiply it by the price at which each product sold, and add up the total. In 2014, the U.S. GDP totaled \$17.4 trillion, the largest GDP in the world.

Each of the market transactions that enter into GDP must involve both a buyer and a seller. The GDP of an economy can be measured either by the total dollar value of what is purchased in the economy, or by the total dollar value of what is produced. There is even a third way, as we will explain later.

GDP Measured by Components of Demand

Who buys all of this production? This demand can be divided into four main parts: consumer spending (consumption), business spending (investment), government spending on goods and services, and spending on net exports. (See the following Clear It Up feature to understand what is meant by investment.) [\[link\]](#) shows how these four components added up to the GDP in 2014. [\[link\]](#) (a) shows the levels of consumption, investment, and

government purchases over time, expressed as a percentage of GDP, while [\[link\]](#) (b) shows the levels of exports and imports as a percentage of GDP over time. A few patterns about each of these components are worth noticing. [\[link\]](#) shows the components of GDP from the demand side.

	Components of GDP on the Demand Side (in trillions of dollars)	Percentage of Total
Consumption	\$11.9	68.4%
Investment	\$2.9	16.7%
Government	\$3.2	18.4%
Exports	\$2.3	13.2%
Imports	−\$2.9	−16.7%
Total GDP	\$17.4	100%

Components of U.S. GDP in 2014: From the Demand Side(Source: http://bea.gov/iTable/index_nipa.cfm)

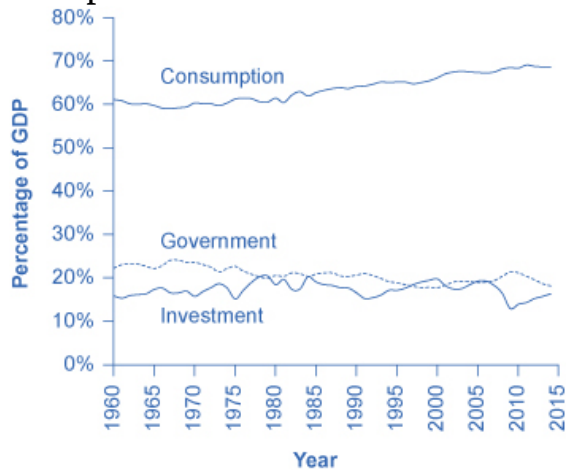
Note:

What is meant by the word “investment”?

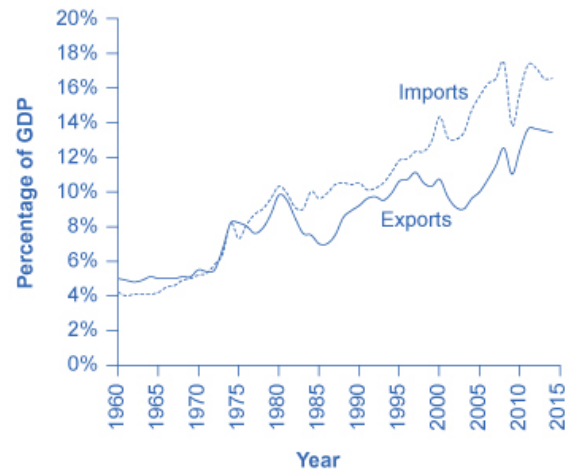
What do economists mean by investment, or business spending? In calculating GDP, investment does not refer to the purchase of stocks and bonds or the trading of financial assets. It refers to the purchase of new capital goods, that is, new commercial real estate (such as buildings, factories, and stores) and equipment, residential housing construction, and

inventories. Inventories that are produced this year are included in this year's GDP—even if they have not yet sold. From the accountant's perspective, it is as if the firm invested in its own inventories. Business investment in 2014 was almost \$3 trillion, according to the Bureau of Economic Analysis.

Components of GDP on the Demand Side



(a) Demand from consumption, investment, and government



(b) Imports and exports

(a) Consumption is about two-thirds of GDP, but it moves relatively little over time. Business investment hovers around 15% of GDP, but it increases and declines more than consumption. Government spending on goods and services is around 20% of GDP. (b) Exports are added to total demand for goods and services, while imports are subtracted from total demand. If exports exceed imports, as in most of the 1960s and 1970s in the U.S. economy, a trade surplus exists. If imports exceed exports, as in recent years, then a trade deficit exists. (Source: http://bea.gov/iTable/index_nipa.cfm)

Consumption expenditure by households is the largest component of GDP, accounting for about two-thirds of the GDP in any year. This tells us that consumers' spending decisions are a major driver of the economy. However,

consumer spending is a gentle elephant: when viewed over time, it does not jump around too much.

Investment expenditure refers to purchases of physical plant and equipment, primarily by businesses. If Starbucks builds a new store, or Amazon buys robots, these expenditures are counted under business investment.

Investment demand is far smaller than consumption demand, typically accounting for only about 15–18% of GDP, but it is very important for the economy because this is where jobs are created. However, it fluctuates more noticeably than consumption. Business investment is volatile; new technology or a new product can spur business investment, but then confidence can drop and business investment can pull back sharply.

If you have noticed any of the infrastructure projects (new bridges, highways, airports) launched during the recession of 2009, you have seen how important government spending can be for the economy. Government expenditure in the United States is about 20% of GDP, and includes spending by all three levels of government: federal, state, and local. The only part of government spending counted in demand is government purchases of goods or services produced in the economy. Examples include the government buying a new fighter jet for the Air Force (federal government spending), building a new highway (state government spending), or a new school (local government spending). A significant portion of government budgets are transfer payments, like unemployment benefits, veteran's benefits, and Social Security payments to retirees. These payments are excluded from GDP because the government does not receive a new good or service in return or exchange. Instead they are transfers of income from taxpayers to others. If you are curious about the awesome undertaking of adding up GDP, read the following Clear It Up feature.

When thinking about the demand for domestically produced goods in a global economy, it is important to count spending on exports—domestically produced goods that are sold abroad. By the same token, we must also subtract spending on imports—goods produced in other countries that are purchased by residents of this country. The net export component of GDP is equal to the dollar value of exports (X) minus the dollar value of imports (M), $(X - M)$. The gap between exports and imports is called the **trade balance**. If a country's exports are larger than its imports, then a country is

said to have a **trade surplus**. In the United States, exports typically exceeded imports in the 1960s and 1970s, as shown in [\[link\]](#) (b).

Since the early 1980s, imports have typically exceeded exports, and so the United States has experienced a **trade deficit** in most years. Indeed, the trade deficit grew quite large in the late 1990s and in the mid-2000s. [\[link\]](#) (b) also shows that imports and exports have both risen substantially in recent decades, even after the declines during the Great Recession between 2008 and 2009. As noted before, if exports and imports are equal, foreign trade has no effect on total GDP. However, even if exports and imports are balanced overall, foreign trade might still have powerful effects on particular industries and workers by causing nations to shift workers and physical capital investment toward one industry rather than another.

Based on these four components of demand, GDP can be measured as:

Equation:

GDP = Consumption + Investment + Government + Trade balance

$GDP = C + I + G + (X - M)$

Understanding how to measure GDP is important for analyzing connections in the macro economy and for thinking about macroeconomic policy tools.

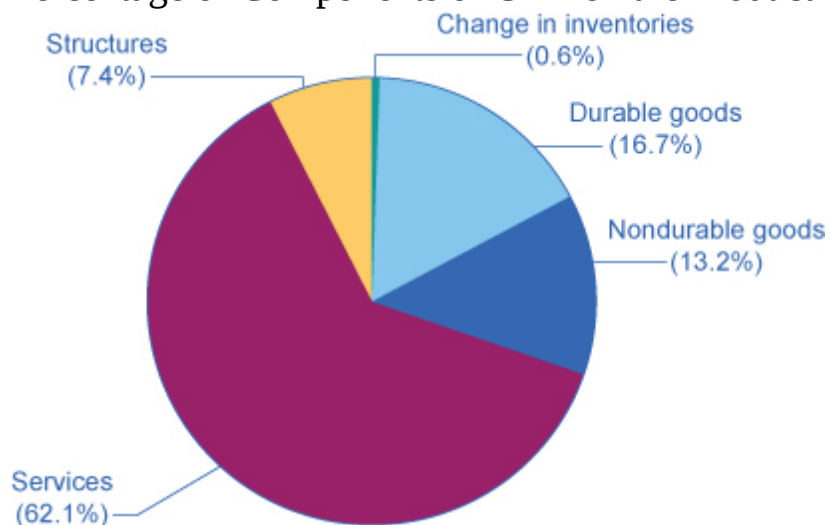
GDP Measured by What is Produced

Everything that is purchased must be produced first. [\[link\]](#) breaks down what is produced into five categories: **durable goods**, **nondurable goods**, **services**, **structures**, and the change in **inventories**. Before going into detail about these categories, notice that total GDP measured according to what is produced is exactly the same as the GDP measured by looking at the five components of demand. [\[link\]](#) provides a visual representation of this information.

	Components of GDP on the Supply Side (in trillions of dollars)	Percentage of Total
Goods		
Durable goods	\$2.9	16.7%
Nondurable goods	\$2.3	13.2%
Services	\$10.8	62.1%
Structures	\$1.3	7.4%
Change in inventories	\$0.1	0.6%
Total GDP	\$17.4	100%

Components of U.S. GDP on the Production Side, 2014(Source: http://bea.gov/iTable/index_nipa.cfm)

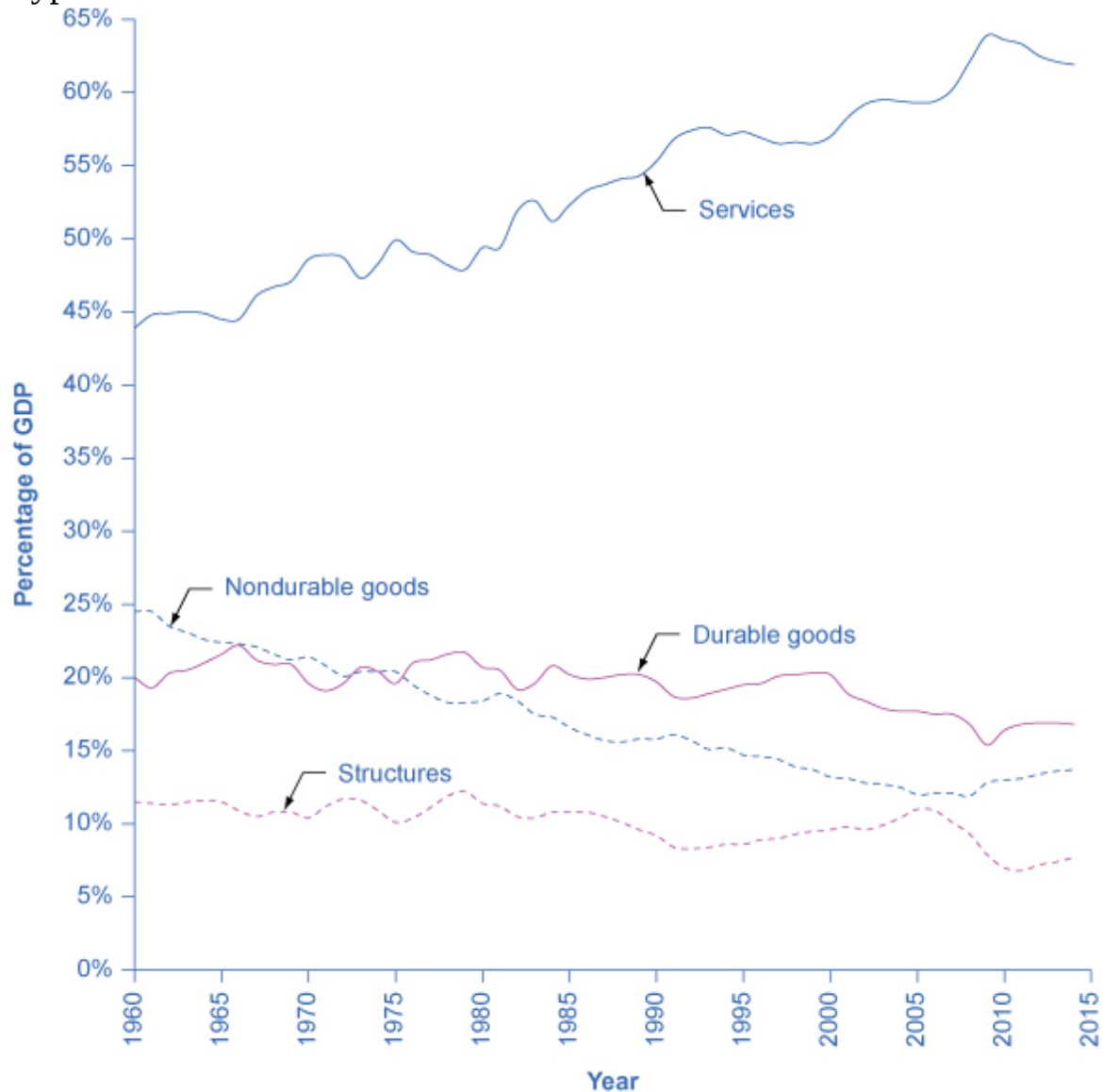
Percentage of Components of GDP on the Production Side



Services make up over half of the production side components of GDP in the United States.

Since every market transaction must have both a buyer and a seller, GDP must be the same whether measured by what is demanded or by what is produced. [\[link\]](#) shows these components of what is produced, expressed as a percentage of GDP, since 1960.

Types of Production



Services are the largest single component of total supply,

representing over half of GDP. Nondurable goods used to be larger than durable goods, but in recent years, nondurable goods have been dropping closer to durable goods, which is about 20% of GDP. Structures hover around 10% of GDP. The change in inventories, the final component of aggregate supply, is not shown here; it is typically less than 1% of GDP.

In thinking about what is produced in the economy, many non-economists immediately focus on solid, long-lasting goods, like cars and computers. By far the largest part of GDP, however, is services. Moreover, services have been a growing share of GDP over time. A detailed breakdown of the leading service industries would include healthcare, education, and legal and financial services. It has been decades since most of the U.S. economy involved making solid objects. Instead, the most common jobs in a modern economy involve a worker looking at pieces of paper or a computer screen; meeting with co-workers, customers, or suppliers; or making phone calls.

Even within the overall category of goods, long-lasting durable goods like cars and refrigerators are about the same share of the economy as short-lived nondurable goods like food and clothing. The category of structures includes everything from homes, to office buildings, shopping malls, and factories. Inventories is a small category that refers to the goods that have been produced by one business but have not yet been sold to consumers, and are still sitting in warehouses and on shelves. The amount of inventories sitting on shelves tends to decline if business is better than expected, or to rise if business is worse than expected.

The Problem of Double Counting

GDP is defined as the current value of all final goods and services produced in a nation in a year. What are final goods? They are goods at the furthest stage of production at the end of a year. Statisticians who calculate GDP must avoid the mistake of **double counting**, in which output is counted more than once as it travels through the stages of production. For example, imagine what would happen if government statisticians first counted the

value of tires produced by a tire manufacturer, and then counted the value of a new truck sold by an automaker that contains those tires. In this example, the value of the tires would have been counted twice—because the price of the truck includes the value of the tires.

To avoid this problem, which would overstate the size of the economy considerably, government statisticians count just the value of **final goods and services** in the chain of production that are sold for consumption, investment, government, and trade purposes. **Intermediate goods**, which are goods that go into the production of other goods, are excluded from GDP calculations. From the example above, only the value of the Ford truck will be counted. The value of what businesses provide to other businesses is captured in the final products at the end of the production chain.

The concept of GDP is fairly straightforward: it is just the dollar value of all final goods and services produced in the economy in a year. In our decentralized, market-oriented economy, actually calculating the more than \$16 trillion-dollar U.S. GDP—along with how it is changing every few months—is a full-time job for a brigade of government statisticians.

What is Counted in GDP	What is not included in GDP
Consumption	Intermediate goods
Business investment	Transfer payments and non-market activities
Government spending on goods and services	Used goods
Net exports	Illegal goods

Counting GDP

Notice the items that are not counted into GDP, as outlined in [\[link\]](#). The sales of used goods are not included because they were produced in a previous year and are part of that year's GDP. The entire underground economy of services paid "under the table" and illegal sales should be counted, but is not, because it is impossible to track these sales. In a recent study by Friedrich Schneider of shadow economies, the underground economy in the United States was estimated to be 6.6% of GDP, or close to \$2 trillion dollars in 2013 alone. Transfer payments, such as payment by the government to individuals, are not included, because they do not represent production. Also, production of some goods—such as home production as when you make your breakfast—is not counted because these goods are not sold in the marketplace.

Other Ways to Measure the Economy

Besides GDP, there are several different but closely related ways of measuring the size of the economy. We mentioned above that GDP can be thought of as total production and as total purchases. It can also be thought of as total income since anything produced and sold produces income.

One of the closest cousins of GDP is the **gross national product (GNP)**. GDP includes only what is produced within a country's borders. GNP adds what is produced by domestic businesses and labor abroad, and subtracts out any payments sent home to other countries by foreign labor and businesses located in the United States. In other words, GNP is based more on the production of citizens and firms of a country, wherever they are located, and GDP is based on what happens within the geographic boundaries of a certain country. For the United States, the gap between GDP and GNP is relatively small; in recent years, only about 0.2%. For small nations, which may have a substantial share of their population working abroad and sending money back home, the difference can be substantial.

Net national product (NNP) is calculated by taking GNP and then subtracting the value of how much physical capital is worn out, or reduced in value because of aging, over the course of a year. The process by which capital ages and loses value is called **depreciation**. The NNP can be further subdivided into **national income**, which includes all income to businesses

and individuals, and personal income, which includes only income to people.

For practical purposes, it is not vital to memorize these definitions. However, it is important to be aware that these differences exist and to know what statistic you are looking at, so that you do not accidentally compare, say, GDP in one year or for one country with GNP or NNP in another year or another country. To get an idea of how these calculations work, follow the steps in the following Work It Out feature.

Key Concepts and Summary

The size of a nation's economy is commonly expressed as its gross domestic product (GDP), which measures the value of the output of all goods and services produced within the country in a year. GDP is measured by taking the quantities of all goods and services produced, multiplying them by their prices, and summing the total. Since GDP measures what is bought and sold in the economy, it can be measured either by the sum of what is purchased in the economy or what is produced.

Demand can be divided into consumption, investment, government, exports, and imports. What is produced in the economy can be divided into durable goods, nondurable goods, services, structures, and inventories. To avoid double counting, GDP counts only final output of goods and services, not the production of intermediate goods or the value of labor in the chain of production.

Self-Check Questions

Exercise:

Problem:

Country A has export sales of \$20 billion, government purchases of \$1,000 billion, business investment is \$50 billion, imports are \$40 billion, and consumption spending is \$2,000 billion. What is the dollar value of GDP?

Solution:

GDP is $C + I + G + (X - M)$. $GDP = \$2,000 \text{ billion} + \$50 \text{ billion} + \$1,000 \text{ billion} + (\$20 \text{ billion} - \$40 \text{ billion}) = \$3,030$

Exercise:**Problem:**

Which of the following are included in GDP, and which are not?

- a. The cost of hospital stays
- b. The rise in life expectancy over time
- c. Child care provided by a licensed day care center
- d. Child care provided by a grandmother
- e. The sale of a used car
- f. The sale of a new car
- g. The greater variety of cheese available in supermarkets
- h. The iron that goes into the steel that goes into a refrigerator bought by a consumer.

Solution:

- a. Hospital stays are part of GDP.
- b. Changes in life expectancy are not market transactions and not part of GDP.
- c. Child care that is paid for is part of GDP.
- d. If Grandma gets paid and reports this as income, it is part of GDP, otherwise not.
- e. A used car is not produced this year, so it is not part of GDP.
- f. A new car is part of GDP.
- g. Variety does not count in GDP, where the cheese could all be cheddar.
- h. The iron is not counted because it is an intermediate good.

Review Questions

Exercise:

Problem:

What are the main components of measuring GDP with what is demanded?

Exercise:

Problem:

What are the main components of measuring GDP with what is produced?

Exercise:

Problem:

Would you usually expect GDP as measured by what is demanded to be greater than GDP measured by what is supplied, or the reverse?

Exercise:

Problem:

Why must double counting be avoided when measuring GDP?

Critical Thinking Question

Exercise:

Problem:

U.S. macroeconomic data are thought to be among the best in the world. Given what you learned in the [Clear It Up](#) "How do statisticians measure GDP?", does this surprise you? Or does this simply reflect the complexity of a modern economy?

Exercise:

Problem: What does GDP not tell us about the economy?

Problem

Exercise:

Problem:

Last year, a small nation with abundant forests cut down \$200 worth of trees. \$100 worth of trees were then turned into \$150 worth of lumber. \$100 worth of that lumber was used to produce \$250 worth of bookshelves. Assuming the country produces no other outputs, and there are no other inputs used in the production of trees, lumber, and bookshelves, what is this nation's GDP? In other words, what is the value of the final goods produced including trees, lumber and bookshelves?

References

U.S. Department of Commerce: Bureau of Economic Analysis. "National data: National Income and Product Accounts Tables."
<http://bea.gov/iTable/iTable.cfm?ReqID=9&step=1>.

U.S. Department of Commerce: United States Census Bureau. "Census of Governments: 2012 Census of Governments."
<http://www.census.gov/govs/cog/>.

United States Department of Transportation. "About DOT." Last modified March 2, 2012. <http://www.dot.gov/about>.

U.S. Department of Energy. "Energy.gov." <http://energy.gov/>.

U.S. Department of Health & Human Services. "Agency for Healthcare Research and Quality." <http://www.ahrq.gov/>.

United States Department of Agriculture. "USDA."
<http://www.usda.gov/wps/portal/usda/usdahome>.

Schneider, Friedrich. Department of Economics. "Size and Development of the Shadow Economy of 31 European and 5 other OECD Countries from

2003 to 2013: A Further Decline.” *Johannes Kepler University*. Last modified April 5, 2013.
http://www.econ.jku.at/members/Schneider/files/publications/2013/ShadEcEurope31_Jan2013.pdf.

Glossary

depreciation

the process by which capital ages over time and therefore loses its value

double counting

a potential mistake to be avoided in measuring GDP, in which output is counted more than once as it travels through the stages of production

durable good

long-lasting good like a car or a refrigerator

final good and service

output used directly for consumption, investment, government, and trade purposes; contrast with “intermediate good”

gross domestic product (GDP)

the value of the output of all goods and services produced within a country in a year

gross national product (GNP)

includes what is produced domestically and what is produced by domestic labor and business abroad in a year

intermediate good

output provided to other businesses at an intermediate stage of production, not for final users; contrast with “final good and service”

inventory

good that has been produced, but not yet been sold

national income

includes all income earned: wages, profits, rent, and profit income

net national product (NNP)

GDP minus depreciation

nondurable good

short-lived good like food and clothing

service

product which is intangible (in contrast to goods) such as entertainment, healthcare, or education

structure

building used as residence, factory, office building, retail store, or for other purposes

trade balance

gap between exports and imports

trade deficit

exists when a nation's imports exceed its exports and is calculated as $\text{imports} - \text{exports}$

trade surplus

exists when a nation's exports exceed its imports and is calculated as $\text{exports} - \text{imports}$

Adjusting Nominal Values to Real Values

By the end of this section, you will be able to:

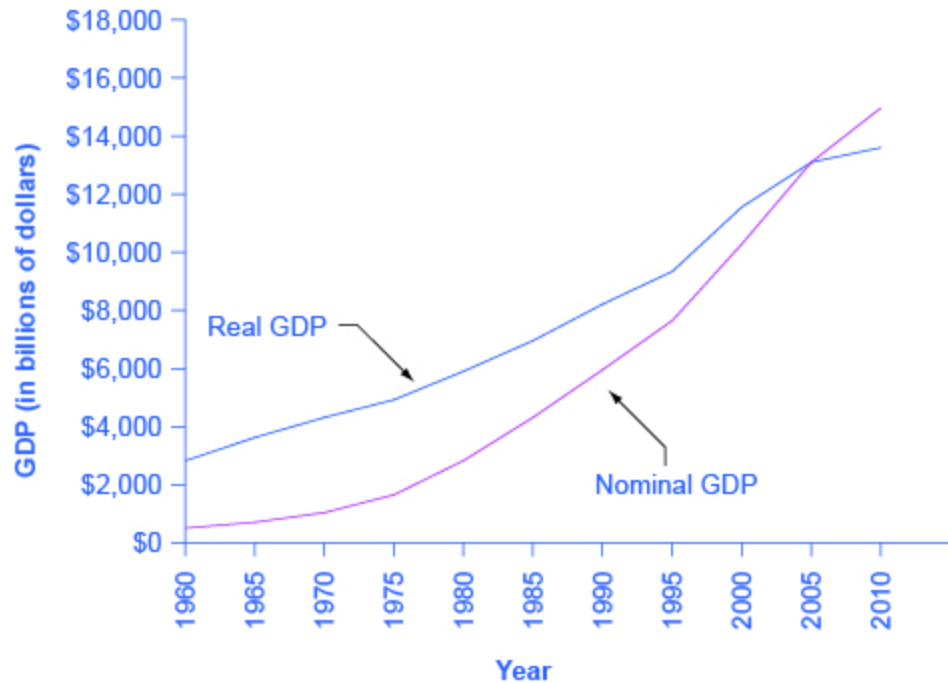
- Contrast nominal GDP and real GDP
- Explain GDP deflator
- Calculate real GDP based on nominal GDP values

When examining economic statistics, there is a crucial distinction worth emphasizing. The distinction is between nominal and real measurements, which refer to whether or not inflation has distorted a given statistic. Looking at economic statistics without considering inflation is like looking through a pair of binoculars and trying to guess how close something is: unless you know how strong the lenses are, you cannot guess the distance very accurately. Similarly, if you do not know the rate of inflation, it is difficult to figure out if a rise in GDP is due mainly to a rise in the overall level of prices or to a rise in quantities of goods produced. The **nominal value** of any economic statistic means the statistic is measured in terms of actual prices that exist at the time. The **real value** refers to the same statistic after it has been adjusted for inflation. Generally, it is the real value that is more important.

Comparing Nominal to Real GDP

If an unwary analyst compared nominal GDP in 1960 to nominal GDP in 2010, it might appear that national output had risen by a factor of twenty-seven over this time (that is, GDP of \$14,958 billion in 2010 divided by GDP of \$543 billion in 1960). This conclusion would be highly misleading. Recall that nominal GDP is defined as the quantity of every good or service produced multiplied by the price at which it was sold, summed up for all goods and services. In order to see how much production has actually increased, we need to extract the effects of higher prices on nominal GDP.

U.S. Nominal and Real GDP, 1960–2012



The red line measures U.S. GDP in nominal dollars. The black line measures U.S. GDP in real dollars, where all dollar values have been converted to 2005 dollars. Since real GDP is expressed in 2005 dollars, the two lines cross in 2005. However, real GDP will appear higher than nominal GDP in the years before 2005, because dollars were worth less in 2005 than in previous years. Conversely, real GDP will appear lower in the years after 2005, because dollars were worth more in 2005 than in later years.

How much did GDP increase in real terms? What was the rate of growth of real GDP from 1960 to 2010? To find the real growth rate, we apply the formula for percentage change:

Equation:

$$\frac{2010 \text{ real GDP} - 1960 \text{ real GDP}}{1960 \text{ real GDP}} \times 100 = \% \text{ change}$$

$$\frac{13,598.5 - 2,859.5}{2,859.5} \times 100 = 376\%$$

In other words, the U.S. economy has increased real production of goods and services by nearly a factor of four since 1960. Of course, that understates the material improvement since it fails to capture improvements in the quality of products and the invention of new products.

Therefore, the growth rate of real GDP (% change in quantity) equals the growth rate in nominal GDP (% change in value) minus the inflation rate (% change in price).

Note that using this equation provides an approximation for small changes in the levels. For more accurate measures, one should use the first formula shown.

Key Concepts and Summary

The nominal value of an economic statistic is the commonly announced value. The real value is the value after adjusting for changes in inflation. To convert nominal economic data from several different years into real, inflation-adjusted data, the starting point is to choose a base year arbitrarily and then use a price index to convert the measurements so that they are measured in the money prevailing in the base year.

Self-Check Question

Exercise:

Problem:

Using data from [\[link\]](#) how much of the nominal GDP growth from 1980 to 1990 was real GDP and how much was inflation?

Solution:

From 1980 to 1990, real GDP grew by $(8,225.0 - 5,926.5) / (5,926.5) = 39\%$. Over the same period, prices increased by $(72.7 - 48.3) / (48.3/100) = 51\%$. So about 57% of the growth $51 / (51 + 39)$ was

inflation, and the remainder: $39 / (51 + 39) = 43\%$ was growth in real GDP.

Review Questions

Exercise:

Problem:

What is the difference between a series of economic data over time measured in nominal terms versus the same data series over time measured in real terms?

Exercise:

Problem:

How do you convert a series of nominal economic data over time to real terms?

Critical Thinking Question

Exercise:

Problem:

Should people typically pay more attention to their real income or their nominal income? If you choose the latter, why would that make sense in today's world? Would your answer be the same for the 1970s?

Problems

Exercise:

Problem:

The “prime” interest rate is the rate that banks charge their best customers. Based on the nominal interest rates and inflation rates given in [\[link\]](#), in which of the years given would it have been best to be a lender? Based on the nominal interest rates and inflation rates given in [\[link\]](#), in which of the years given would it have been best to be a borrower?

Year	Prime Interest Rate	Inflation Rate
1970	7.9%	5.7%
1974	10.8%	11.0%
1978	9.1%	7.6%
1981	18.9%	10.3%

Exercise:**Problem:**

A mortgage loan is a loan that a person makes to purchase a house. [\[link\]](#) provides a list of the mortgage interest rate being charged for several different years and the rate of inflation for each of those years. In which years would it have been better to be a person borrowing money from a bank to buy a home? In which years would it have been better to be a bank lending money?

Year	Mortgage Interest Rate	Inflation Rate
1984	12.4%	4.3%
1990	10%	5.4%
2001	7.0%	2.8%

Glossary

nominal value

the economic statistic actually announced at that time, not adjusted for inflation; contrast with real value

real value

an economic statistic after it has been adjusted for inflation; contrast with nominal value

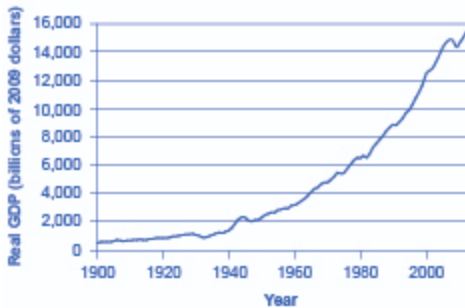
Tracking Real GDP over Time

By the end of this section, you will be able to:

- Explain recessions, depressions, peaks, and troughs
- Evaluate the importance of tracking real GDP over time

When news reports indicate that “the economy grew 1.2% in the first quarter,” the reports are referring to the percentage change in real GDP. By convention, GDP growth is reported at an annualized rate: Whatever the calculated growth in real GDP was for the quarter, it is multiplied by four when it is reported as if the economy were growing at that rate for a full year.

U.S. GDP, 1900–2014



Real GDP in the United States in 2014 was about \$16 trillion. After adjusting to remove the effects of inflation, this represents a roughly 20-fold increase in the economy’s production of goods and services since the start of the twentieth century.
(Source: bea.gov)

[\[link\]](#) shows the pattern of U.S. real GDP since 1900. The generally upward long-term path of GDP has been regularly interrupted by short-term declines. A significant decline in real GDP is called a **recession**. An especially lengthy and deep recession is called a **depression**. The severe drop in GDP that occurred during the Great Depression of the 1930s is clearly visible in the figure, as is the Great Recession of 2008–2009.

Real GDP is important because it is highly correlated with other measures of economic activity, like employment and unemployment. When real GDP rises, so does employment.

The most significant human problem associated with recessions (and their larger, uglier cousins, depressions) is that a slowdown in production means that firms need to lay off or fire some of the workers they have. Losing a job imposes painful financial and personal costs on workers, and often on their extended families as well. In addition, even those who keep their jobs are likely to find that wage raises are scanty at best—or they may even be asked to take pay cuts.

As we consider the changes in the output of an economy over time, it is helpful to remember that the changes occur in cycles. While there are a wide variety of definitions for the term **recession**, the one that most economists agree on is that a recession is characterized by two consecutive quarters (6 months) of decline in GDP. The highest point of the economy, before a recession begins, is called the **peak**; conversely, the lowest point of a recession, before a recovery begins, is called the **trough**. Thus, a recession lasts from peak to trough, and an economic upswing, known as an **expansion or recovery** runs from trough to peak. The movement of the economy from peak to trough and trough to peak is called the **business cycle**.

While economists differ in their definition of depression, most agree that a **depression** is "longer and more severe" than a recession. In the words of one economist, "you'll know it when you see it". The most significant depression in U.S. history is called the Great Depression, and lasted for over a decade.

[\[link\]](#) lists the pattern of recessions and expansions in the U.S. economy since 1900. It is intriguing to notice that the three longest trough-to-peak expansions of the twentieth century have happened since 1960. The most recent recession started in December 2007 and ended formally in June 2009. This was the most severe recession since the Great Depression of the 1930's.

Trough	Peak	Months of Contraction	Months of Expansion
December 1900	September 1902	18	21
August 1904	May 1907	23	33
June 1908	January 1910	13	19
January 1912	January 1913	24	12
December 1914	August 1918	23	44
March 1919	January 1920	7	10
July 1921	May 1923	18	22
July 1924	October 1926	14	27

Trough	Peak	Months of Contraction	Months of Expansion
November 1927	August 1929	23	21
March 1933	May 1937	43	50
June 1938	February 1945	13	80
October 1945	November 1948	8	37
October 1949	July 1953	11	45
May 1954	August 1957	10	39
April 1958	April 1960	8	24
February 1961	December 1969	10	106
November 1970	November 1973	11	36
March 1975	January 1980	16	58
July 1980	July 1981	6	12

Trough	Peak	Months of Contraction	Months of Expansion
November 1982	July 1990	16	92
March 2001	November 2001	8	120
December 2007	June 2009	18	73

U.S. Business Cycles since 1900(Source:
<http://www.nber.org/cycles/main.html>)

A private think tank, the National Bureau of Economic Research (NBER), is the official tracker of business cycles for the U.S. economy. However, the effects of a severe recession often linger on after the official ending date assigned by the NBER.

Key Concepts and Summary

Over the long term, U.S. real GDP have increased dramatically. At the same time, GDP has not increased the same amount each year. The speeding up and slowing down of GDP growth represents the business cycle. When GDP declines significantly, a recession occurs. A longer and deeper decline is a depression. Recessions begin at the peak of the business cycle and end at the trough.

Self-Check Questions

Exercise:

Problem:

Without looking at [\[link\]](#), return to [\[link\]](#). If a recession is defined as a significant decline in national output, can you identify any post-1960 recessions in addition to the recession of 2008–2009? (This requires a judgment call.)

Solution:

Two other major recessions are visible in the figure as slight dips: those of 1973–1975, and 1981–1982. Two other recessions appear in the figure as a flattening of the path of real GDP. These were in 1990–1991 and 2001.

Exercise:**Problem:**

According to [\[link\]](#), how often have recessions occurred since the end of World War II (1945)?

Solution:

11 recessions in approximately 70 years averages about one recession every six years.

Exercise:**Problem:**

According to [\[link\]](#), how long has the average recession lasted since the end of World War II?

Solution:

The table lists the “Months of Contraction” for each recession. Averaging these figures for the post-WWII recessions gives an average duration of 11 months, or slightly less than a year.

Exercise:

Problem:

According to [\[link\]](#), how long has the average expansion lasted since the end of World War II?

Solution:

The table lists the “Months of Expansion.” Averaging these figures for the post-WWII expansions gives an average expansion of 60.5 months, or more than five years.

Review Question**Exercise:****Problem:**

What are the typical patterns of GDP for a high-income economy like the United States in the long run and the short run?

Critical Thinking Questions**Exercise:****Problem:**

Why do you suppose that U.S. GDP is so much higher today than 50 or 100 years ago?

Exercise:**Problem:**

Why do you think that GDP does not grow at a steady rate, but rather speeds up and slows down?

References

The National Bureau of Economic Research. “Information on Recessions and Recoveries, the NBER Business Cycle Dating Committee, and related topics.” <http://www.nber.org/cycles/main.html>.

Glossary

business cycle

the relatively short-term movement of the economy in and out of recession

depression

an especially lengthy and deep decline in output

peak

during the business cycle, the highest point of output before a recession begins

recession

a significant decline in national output

trough

during the business cycle, the lowest point of output in a recession, before a recovery begins

Comparing GDP among Countries

By the end of this section, you will be able to:

- Explain how GDP can be used to compare the economic welfare of different nations
- Calculate the conversion of GDP to a common currency by using exchange rates
- Calculate GDP per capita using population data

It is common to use GDP as a measure of economic welfare or standard of living in a nation. When comparing the GDP of different nations for this purpose, two issues immediately arise. First, the GDP of a country is measured in its own currency: the United States uses the U.S. dollar; Canada, the Canadian dollar; most countries of Western Europe, the euro; Japan, the yen; Mexico, the peso; and so on. Thus, comparing GDP between two countries requires converting to a common currency. A second issue is that countries have very different numbers of people. For instance, the United States has a much larger economy than Mexico or Canada, but it also has roughly three times as many people as Mexico and nine times as many people as Canada. So, if we are trying to compare standards of living across countries, we need to divide GDP by population.

GDP Per Capita

The U.S. economy has the largest GDP in the world, by a considerable amount. The United States is also a populous country; in fact, it is the third largest country by population in the world, although well behind China and India. So is the U.S. economy larger than other countries just because the United States has more people than most other countries, or because the U.S. economy is actually larger on a per-person basis? This question can be answered by calculating a country's **GDP per capita**; that is, the GDP divided by the population.

Equation:

Country	GDP (in billions of U.S. dollars)	Population (in millions)	Per Capita GDP (in U.S. dollars)
Brazil	2,246.00	199.20	11,172.50
Canada	1,826.80	35.10	52,037.10
China	9,469.10	1,360.80	6,958.70
Egypt	271.40	83.70	3,242.90
Germany	3,636.00	80.80	44,999.50
India	1,876.80	1,243.30	1,509.50
Japan	4,898.50	127.3	38,467.80
Mexico	1,260.90	118.40	10,649.90
South Korea	1,304.47	50.20	25,975.10
United Kingdom	2,523.20	64.10	39,371.70
United States	16,768.10	316.30	53,001.00

GDP Per Capita, 2013(Source:
<http://www.imf.org/external/pubs/ft/weo/2013/01/weodata/index.aspx>)

Notice that the ranking by GDP is different from the ranking by GDP per capita. India has a somewhat larger GDP than Germany, but on a per capita basis, Germany has more than 10 times India's standard of living. Will China soon have a better standard of living than the U.S.? Read the following Clear It Up feature to find out.

The high-income nations of the world—including the United States, Canada, the Western European countries, and Japan—typically have GDP per capita in the range of \$20,000 to \$50,000. Middle-income countries, which include much of Latin America, Eastern Europe, and some countries in East Asia, have GDP per capita in the range of \$6,000 to \$12,000. The low-income countries in the world, many of them located in Africa and Asia, often have GDP per capita of less than \$2,000 per year.

Key Concepts and Summary

Since GDP is measured in a country's currency, in order to compare different countries' GDPs, we need to convert them to a common currency. One way to do that is with the exchange rate, which is the price of one country's currency in terms of another. Once GDPs are expressed in a common currency, we can compare each country's GDP per capita by dividing GDP by population. Countries with large populations often have large GDPs, but GDP alone can be a misleading indicator of the wealth of a nation. A better measure is GDP per capita.

Self-Check Question

Exercise:

Problem:

Is it possible for GDP to rise while at the same time per capita GDP is falling? Is it possible for GDP to fall while per capita GDP is rising?

Solution:

Yes. The answer to both questions depends on whether GDP is growing faster or slower than population. If population grows faster than GDP, GDP increases, while GDP per capita decreases. If GDP falls, but population falls faster, then GDP decreases, while GDP per capita increases.

Exercise:

Problem:

The Central African Republic has a GDP of 1,107,689 million CFA francs and a population of 4.862 million. The exchange rate is 284.681 CFA francs per dollar. Calculate the GDP per capita of Central African Republic.

Solution:

Start with Central African Republic's GDP measured in francs. Divide it by the exchange rate to convert to U.S. dollars, and then divide by population to obtain the per capita figure. That is, $1,107,689 \text{ million francs} / 284.681 \text{ francs per dollar} / 4.862 \text{ million people} = \800.28 GDP per capita.

Review Question**Exercise:****Problem:**

What are the two main difficulties that arise in comparing the GDP of different countries?

Critical Thinking Question**Exercise:****Problem:**

Cross country comparisons of GDP per capita typically use purchasing power parity equivalent exchange rates, which are a measure of the long run equilibrium value of an exchange rate. In fact, we used PPP equivalent exchange rates in this module. Why could using market exchange rates, which sometimes change dramatically in a short period of time, be misleading?

Exercise:**Problem:**

Why might per capita GDP be only an imperfect measure of a country's standard of living?

Problems**Exercise:****Problem:**

Ethiopia has a GDP of \$8 billion (measured in U.S. dollars) and a population of 55 million. Costa Rica has a GDP of \$9 billion (measured in U.S. dollars) and a population of 4 million. Calculate the per capita GDP for each country and identify which one is higher.

Exercise:**Problem:**

In 1980, Denmark had a GDP of \$70 billion (measured in U.S. dollars) and a population of 5.1 million. In 2000, Denmark had a GDP of \$160 billion (measured in U.S. dollars) and a population of 5.3 million. By what percentage did Denmark's GDP per capita rise between 1980 and 2000?

Exercise:**Problem:**

The Czech Republic has a GDP of 1,800 billion koruny. The exchange rate is 20 koruny/U.S. dollar. The Czech population is 20 million. What is the GDP per capita of the Czech Republic expressed in U.S. dollars?

Glossary

exchange rate

the price of one currency in terms of another currency

GDP per capita

GDP divided by the population

Introduction to Unemployment
class="introduction"
Out of Business

Borders was
one of the
many
companies
unable to
recover from
the economic
recession of
2008-2009.

(Credit:
modification
of work by
Luis Villa del
Campo/Flick
r Creative
Commons)



Note:

The Mysterious Case of the Missing Candidates

Nearly eight million U.S. jobs were lost during the Great Recession of 2008-2009, with unemployment peaking at 10% in October 2009, according to the Bureau of Labor Statistics (BLS). That is a huge number of positions gone. During the tepid recovery, some positions were added, but as of summer 2013, unemployment had remained persistently higher than the pre-recession rate of less than 5%. Some economists and policymakers worried the recovery would be “jobless.” With the economy growing, albeit slowly, why wasn’t the unemployment number falling? Why were firms not hiring?

Peter Cappelli, noted Wharton management professor and Director of Wharton’s Center for Human Resources, does not believe the job search process is akin to what he terms the “Home Depot” view of hiring. According to him, this view “basically says that filling a job is like replacing a part in a washing machine. You simply find someone who does the exact same job as that broken part, plug him or her into the washing

machine and that is it.” The job search, for both the prospective employee and the employer, is more complex than that.

In a hiring situation, employers hold all the cards. They write the job descriptions, determine the salaries, decide when and how to advertise positions, and set the controls on employment application screening software. Advertising for positions has increased as the economic recovery progresses, yet here’s the kicker: Employers say there are no applicants out there who meet their needs. While the unemployment rate is now below 6% as of the beginning of 2015, many economists and policymakers (including the Chair of the Federal Reserve, Janet Yellen) are still concerned about "slack" in the labor market. So the question arises: where are the job candidates?

That question leads us to the topic of this chapter—unemployment. What constitutes it? How is it measured? And if the economy is growing, why isn’t the pool of job openings growing along with it? Sounds like the economy has a case of “missing” candidates.

Note:**Introduction to Unemployment**

In this chapter, you will learn about:

- How the Unemployment Rate is Defined and Computed
- Patterns of Unemployment
- What Causes Changes in Unemployment over the Short Run
- What Causes Changes in Unemployment over the Long Run

Unemployment can be a terrible and wrenching life experience—like a serious automobile accident or a messy divorce—whose consequences can be fully understood only by someone who has gone through it. For unemployed individuals and their families, there is the day-to-day financial stress of not knowing where the next paycheck is coming from. There are painful adjustments, like watching your savings account dwindle, selling a car and buying a cheaper one, or moving to a less expensive place to live.

Even when the unemployed person finds a new job, it may pay less than the previous one. For many people, their job is an important part of their self worth. When unemployment separates people from the workforce, it can affect family relationships as well as mental and physical health.

The human costs of unemployment alone would justify making a low level of unemployment an important public policy priority. But unemployment also includes economic costs to the broader society. When millions of unemployed but willing workers cannot find jobs, an economic resource is going unused. An economy with high unemployment is like a company operating with a functional but unused factory. The opportunity cost of unemployment is the output that could have been produced by the unemployed workers.

This chapter will discuss how the unemployment rate is defined and computed. It will examine the patterns of unemployment over time, for the U.S. economy as a whole, for different demographic groups in the U.S. economy, and for other countries. It will then consider an economic explanation for unemployment, and how it explains the patterns of unemployment and suggests public policies for reducing it.

How the Unemployment Rate is Defined and Computed

By the end of this section, you will be able to:

- Calculate the labor force percentage and the unemployment rate
- Explain hidden unemployment and what it means to be in or out of the labor force
- Evaluate the collection and interpretation of unemployment data

Unemployment is typically described in newspaper or television reports as a percentage or a rate. A recent report might have said, for example, *from August 2009 to November 2009, the U.S. unemployment rate rose from 9.7% to 10.0%, but by June 2010, it had fallen to 9.5%*. At a glance, the changes between the percentages may seem small. But remember that the U.S. economy has about 155 million adults who either have jobs or are looking for them. A rise or fall of just 0.1% in the unemployment rate of 155 million potential workers translates into 155,000 people, which is roughly the total population of a city like Syracuse, New York, Brownsville, Texas, or Pasadena, California. Large rises in the unemployment rate mean large numbers of job losses. In November 2009, at the peak of the recession, about 15 million people were out of work. Even with the unemployment rate now at 5.5% as of February 2015, about 8 million people total are out of work.

Who's In or Out of the Labor Force?

Should everyone without a job be counted as unemployed? Of course not. Children, for example, should not be counted as unemployed. Surely, the retired should not be counted as unemployed. Many full-time college students have only a part-time job, or no job at all, but it seems inappropriate to count them as suffering the pains of unemployment. Some people are not working because they are rearing children, ill, on vacation, or on parental leave.

The point is that the adult population is not just divided into employed and unemployed. A third group exists: people who do not have a job, and for some reason—retirement, looking after children, taking a voluntary break

before a new job—are not interested in having a job, either. It also includes those who do want a job but have quit looking, often due to being discouraged by their inability to find suitable employment. Economists refer to this third group of those who are not working and not looking for work as **out of the labor force** or not in the labor force.

The U.S. unemployment rate, which is based on a monthly survey carried out by the U.S. Bureau of the Census, asks a series of questions to divide up the adult population (16 years of age or older) into employed, unemployed, or not in the labor force. To be classified as unemployed, an adult person must be without a job, currently available to work, and actively looking for work in the previous four weeks. Thus, a person who does not have a job but who is not currently available to work or has not actively looked for work in the last four weeks is counted as out of the labor force.

Employed: currently working for pay

Unemployed: Out of work and actively looking for a job

Out of the labor force: Out of paid work and not actively looking for a job

Labor force: the number of employed plus the unemployed

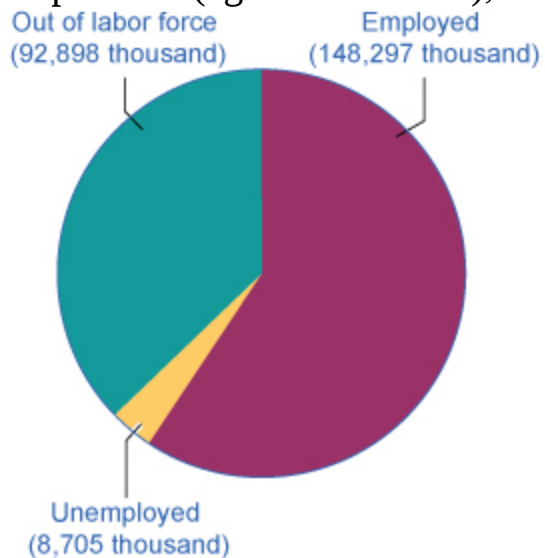
Calculating the Unemployment Rate

[\[link\]](#) shows the three-way division of the over-16 adult population. In February 2015, about 62.8% of the adult population was "in the labor force"; that is, people are either employed or without a job but looking for work. Those in the labor force can be divided into the employed and the unemployed. These values are also shown in [\[link\]](#). The **unemployment rate** is not the percentage of the total adult population without jobs, but rather the percentage of adults who are in the labor force but who do not have jobs:

Equation:

$$\text{Unemployment rate} = \frac{\text{Unemployed people}}{\text{Total labor force}} \times 100$$

Employed, Unemployed, and Out of the Labor Force Distribution of Adult Population (age 16 and older), February 2015



The total adult, working-age population in February 2015 was 249.9 million. Out of this total population, 148.3 were classified as employed, and 8.7 million were classified as unemployed. The remaining 92.9 were classified as out of the labor force. As you will learn, however, this seemingly simple chart does not tell the whole story.



Total adult population over the age of 16	249.9 million
In the labor force	157 million (62.8%)
Employed	148.3 million
Unemployed	8.7 million
Out of the labor force	92.9 million (37.2%)

U.S. Employment and Unemployment, February 2015(Source: <http://www.bls.gov/news.release/empsit.t01.htm>)

In this example, the unemployment rate can be calculated as 8.7 million unemployed people divided by 157 million people in the labor force, which works out to a 5.5% rate of unemployment. The following Work It Out feature will walk you through the steps of this calculation.

Hidden Unemployment

Even with the “out of the labor force” category, there are still some people that are mislabeled in the categorization of employed, unemployed, or out of the labor force. There are some people who have only part time or temporary jobs and who are looking for full time and permanent employment that are counted as employed, though they are not employed in the way they would like or need to be. Additionally, there are individuals who are **underemployed**. This includes those that are trained or skilled for one type or level of work who are working in a lower paying job or one that does not utilize their skills. For example, an individual with a college degree in finance who is working as a sales clerk would be considered underemployed. They are, however, also counted in the employed group. All of these individuals fall under the umbrella of the term “hidden unemployment.” **Discouraged workers**, those who have stopped looking for employment and, hence, are no longer counted in the unemployed also fall into this group

Labor Force Participation Rate

Another important statistic is the **labor force participation rate**. This is the percentage of adults in an economy who are either employed or who are unemployed and looking for a job. So, using the data in [\[link\]](#) and [\[link\]](#), those included in this calculation would be the 157 million individuals in the labor force. The rate is calculated by taking the number of people in the labor force, that is, the number employed and the number unemployed, divided by the total adult population and multiplying by 100 to get the percentage. For the data from February 2015, the labor force participation rate is 62.8%. Historically, the civilian labor force participation rate in the United States climbed beginning in the 1960s as women increasingly entered the workforce, and it peaked at around 68% in late 1999 to early 2000. Since then, the labor force participation rate has steadily declined.

The Establishment Payroll Survey

When the unemployment report comes out each month, the Bureau of Labor Statistics (BLS) also reports on the number of jobs created—which comes from the establishment payroll survey. The payroll survey is based on a survey of about 140,000 businesses and government agencies throughout the United States. It generates payroll employment estimates by the following criteria: all employees, average weekly hours worked, and average hourly, weekly, and overtime earnings. One of the criticisms of this survey is that it does not count the self-employed. It also does not make a distinction between new, minimum wage, part time or temporary jobs and full time jobs with “decent” pay.

How Is the U.S. Unemployment Data Collected?

The unemployment rate announced by the U.S. Bureau of Labor Statistics each month is based on the Current Population Survey (CPS), which has been carried out every month since 1940. Great care is taken to make this survey representative of the country as a whole. The country is first divided into 3,137 areas. The U.S. Bureau of the Census then selects 729 of these areas to survey. The 729 areas are then divided into districts of about 300

households each, and each district is divided into clusters of about four dwelling units. Every month, Census Bureau employees call about 15,000 of the four-household clusters, for a total of 60,000 households. Households are interviewed for four consecutive months, then rotated out of the survey for eight months, and then interviewed again for the same four months the following year, before leaving the sample permanently.

Based on this survey, unemployment rates are calculated by state, industry, urban and rural areas, gender, age, race or ethnicity, and level of education. A wide variety of other information is available, too. For example, how long have people been unemployed? Did they become unemployed because they quit, or were laid off, or their employer went out of business? Is the unemployed person the only wage earner in the family?

Criticisms of Measuring Unemployment

There are always complications in measuring the number of unemployed. For example, what about people who do not have jobs and would be available to work, but have gotten discouraged at the lack of available jobs in their area and stopped looking? Such people, and their families, may be suffering the pains of unemployment. But the survey counts them as out of the labor force because they are not actively looking for work. Other people may tell the Census Bureau that they are ready to work and looking for a job but, truly, they are not that eager to work and are not looking very hard at all. They are counted as unemployed, although they might more accurately be classified as out of the labor force. Still other people may have a job, perhaps doing something like yard work, child care, or cleaning houses, but are not reporting the income earned to the tax authorities. They may report being unemployed, when they actually are working.

Although the unemployment rate gets most of the public and media attention, economic researchers at the Bureau of Labor Statistics publish a wide array of surveys and reports that try to measure these kinds of issues and to develop a more nuanced and complete view of the labor market. It is not exactly a hot news flash that economic statistics are imperfect. Even imperfect measures like the unemployment rate, however, can still be quite informative, when interpreted knowledgeably and sensibly.

Key Concepts and Summary

Unemployment imposes high costs. Unemployed individuals suffer from loss of income and from stress. An economy with high unemployment suffers an opportunity cost of unused resources. The adult population can be divided into those in the labor force and those out of the labor force. In turn, those in the labor force are divided into employed and unemployed. A person without a job must be willing and able to work and actively looking for work to be counted as unemployed; otherwise, a person without a job is counted as being out of the labor force. The unemployment rate is defined as the number of unemployed persons divided by the number of persons in the labor force (not the overall adult population). The Current Population Survey (CPS) conducted by the United States Census Bureau measures the percentage of the labor force that is unemployed. The establishment payroll survey by the Bureau of Labor Statistics measures the net change in jobs created for the month.

Self-Check Questions

Exercise:

Problem:

Suppose the adult population over the age of 16 is 237.8 million and the labor force is 153.9 million (of whom 139.1 million are employed). How many people are “not in the labor force?” What are the proportions of employed, unemployed and not in the labor force in the population? *Hint:* Proportions are percentages.

Solution:

The population is divided into those “in the labor force” and those “not in the labor force.” Thus, the number of adults not in the labor force is $237.8 - 153.9 = 83.9$ million. Since the labor force is divided into employed persons and unemployed persons, the number of unemployed persons is $153.9 - 139.1 = 14.8$ million. Thus, the adult population has the following proportions:

- $139.1/237.8 = 58.5\%$ employed persons
- $14.8/237.8 = 6.2\%$ unemployed persons
- $83.9/237.8 = 35.3\%$ persons out of the labor force

Exercise:

Problem:

Using the above data, what is the unemployment rate? These data are U.S. statistics from 2010. How does it compare to the February 2015 unemployment rate computed earlier?

Solution:

The unemployment rate is defined as the number of unemployed persons as a percentage of the labor force or $14.8/153.9 = 9.6\%$. This is higher than the February 2015 unemployment rate, computed earlier, of 5.5%.

Review Questions

Exercise:

Problem:

What is the difference between being unemployed and being out of the labor force?

Exercise:

Problem:

How is the unemployment rate calculated? How is the labor force participation rate calculated?

Exercise:

Problem: Are all adults who do not hold jobs counted as unemployed?

Exercise:

Problem:

If you are out of school but working part time, are you considered employed or unemployed in U.S. labor statistics? If you are a full time student and working 12 hours a week at the college cafeteria are you considered employed or not in the labor force? If you are a senior citizen who is collecting social security and a pension and working as a greeter at Wal-Mart are you considered employed or not in the labor force?

Exercise:**Problem:**

What happens to the unemployment rate when unemployed workers are reclassified as discouraged workers?

Exercise:**Problem:**

What happens to the labor force participation rate when employed individuals are reclassified as unemployed? What happens when they are reclassified as discouraged workers?

Exercise:**Problem:**

What are some of the problems with using the unemployment rate as an accurate measure of overall joblessness?

Exercise:**Problem:**

What criteria are used by the BLS to count someone as employed? As unemployed?

Exercise:

Problem:

Assess whether the following would be counted as “unemployed” in the Current Employment Statistics survey.

- A. A husband willingly stays home with children while his wife works.
- B. A manufacturing worker whose factory just closed down.
- C. A college student doing an unpaid summer internship.
- D. A retiree.
- E. Someone who has been out of work for two years but keeps looking for a job.
- F. Someone who has been out of work for two months but isn't looking for a job.
- G. Someone who hates her present job and is actively looking for another one.
- H. Someone who decides to take a part time job because she could not find a full time position.

Critical Thinking Questions**Exercise:****Problem:**

Using the definition of the unemployment rate, is an increase in the unemployment rate necessarily a bad thing for a nation?

Exercise:**Problem:**

Is a decrease in the unemployment rate necessarily a good thing for a nation? Explain.

Exercise:

Problem:

If many workers become discouraged from looking for jobs, explain how the number of jobs could decline but the unemployment rate could fall at the same time.

Exercise:**Problem:**

Would you expect hidden unemployment to be higher, lower, or about the same when the unemployment rate is high, say 10%, versus low, say 4%? Explain.

Problems**Exercise:****Problem:**

A country with a population of eight million adults has five million employed, 500,000 unemployed, and the rest of the adult population is out of the labor force. What's the unemployment rate? What share of population is in the labor force? Sketch a pie chart that divides the adult population into these three groups.

Glossary

discouraged workers

those who have stopped looking for employment due to the lack of suitable positions available

labor force participation rate

this is the percentage of adults in an economy who are either employed or who are unemployed and looking for a job

out of the labor force

those who are not working and not looking for work—whether they want employment or not; also termed “not in the labor force”

underemployed

individuals who are employed in a job that is below their skills

unemployment rate

the percentage of adults who are in the labor force and thus seeking jobs, but who do not have jobs

Patterns of Unemployment

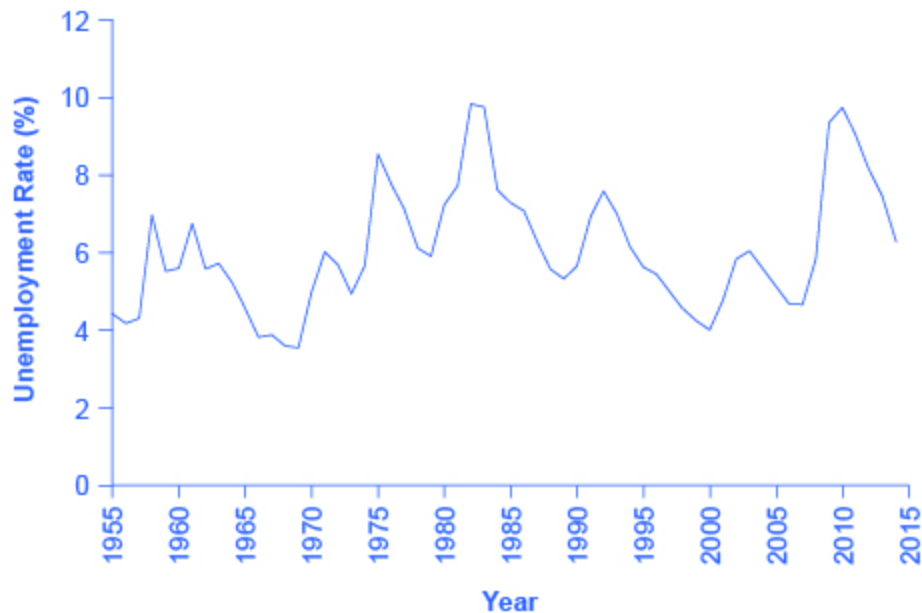
By the end of this section, you will be able to:

- Explain historical patterns of unemployment in the U.S.
- Identify trends of unemployment based on demographics
- Evaluate global unemployment rates

Let's look at how unemployment rates have changed over time and how various groups of people are affected by unemployment differently.

The Historical U.S. Unemployment Rate

[\[link\]](#) shows the historical pattern of U.S. unemployment since 1955. The U.S. Unemployment Rate, 1955–2015



The U.S. unemployment rate moves up and down as the economy moves in and out of recessions. But over time, the unemployment rate seems to return to a range of 4% to 6%. There does not seem to be a long-term trend toward the rate moving generally higher or generally lower. (Source: Federal Reserve Economic Data (FRED) <https://research.stlouisfed.org/fred2/series/LRUN64TTUSA156S0>)

As we look at this data, several patterns stand out:

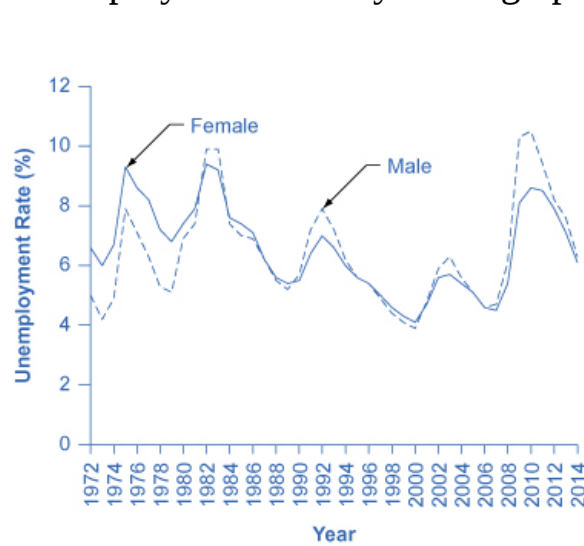
1. Unemployment rates do fluctuate over time. During the deep recessions of the early 1980s and of 2007–2009, unemployment reached roughly 10%. For comparison, during the Great Depression of the 1930s, the unemployment rate reached almost 25% of the labor force.
2. Unemployment rates in the late 1990s and into the mid-2000s were rather low by historical standards. The unemployment rate was below 5% from 1997 to 2000 and near 5% during almost all of 2006–2007. The previous time unemployment had been less than 5% for three consecutive years was three decades earlier, from 1968 to 1970.
3. The unemployment rate never falls all the way to zero. Indeed, it never seems to get below 3%—and it stays that low only for very short periods. (Reasons why this is the case are discussed later in this chapter.)
4. The timing of rises and falls in unemployment matches fairly well with the timing of upswings and downswings in the overall economy. During periods of recession and depression, unemployment is high. During periods of economic growth, unemployment tends to be lower.
5. No significant upward or downward trend in unemployment rates is apparent. This point is especially worth noting because the U.S. population nearly quadrupled from 76 million in 1900 to over 314 million by 2012. Moreover, a higher proportion of U.S. adults are now in the paid workforce, because women have entered the paid labor force in significant numbers in recent decades. Women composed 18% of the paid workforce in 1900 and nearly half of the paid workforce in 2012. But despite the increased number of workers, as well as other economic events like globalization and the continuous invention of new technologies, the economy has provided jobs without causing any long-term upward or downward trend in unemployment rates.

Unemployment Rates by Group

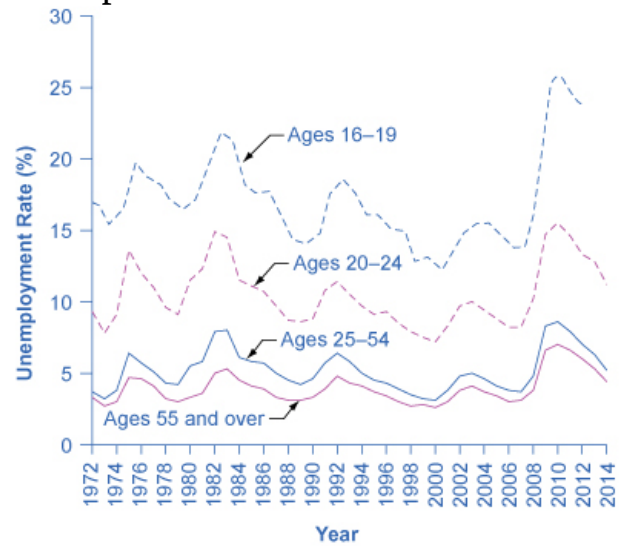
Unemployment is not distributed evenly across the U.S. population. [\[link\]](#) shows unemployment rates broken down in various ways: by gender, age,

and race/ethnicity.

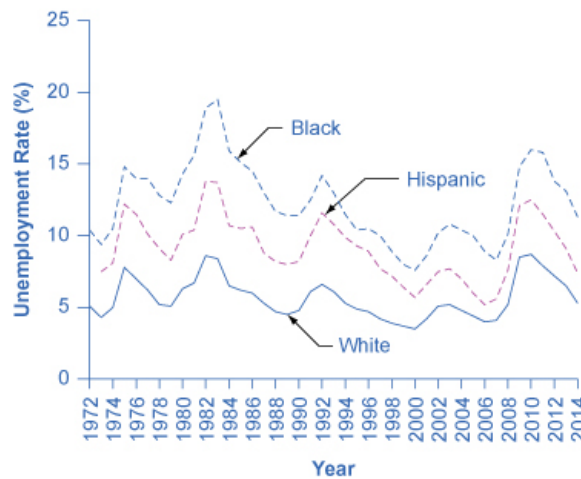
Unemployment Rate by Demographic Group



(a) Unemployment rates by gender



(b) Unemployment rates for women, by age



(c) Unemployment rates by race and ethnicity

(a) By gender, 1972–2014. Unemployment rates for men used to be lower than unemployment rates for women, but in recent decades, the two rates have been very close, often with the unemployment rate for men somewhat higher. (b) By age, 1972–2014. Unemployment rates are highest for the very young and become lower with age. (c) By race and ethnicity, 1972–2014. Although unemployment rates for all groups tend to rise and fall together, the unemployment rate for whites has

been lower than the unemployment rate for blacks and Hispanics in recent decades. (Source: www.bls.gov)

The unemployment rate for women had historically tended to be higher than the unemployment rate for men, perhaps reflecting the historical pattern that women were seen as “secondary” earners. By about 1980, however, the unemployment rate for women was essentially the same as that for men, as shown in [\[link\]](#) (a). During the recession of 2008-2009, the unemployment rate for men exceeded the unemployment rate for women. Through 2014, this pattern has remained, although the gap is narrowing.

Note:

Read this [report](#) for detailed information on the recession of 2008–2009. It also provides some very useful information on the statistics of unemployment.

Younger workers tend to have higher unemployment, while middle-aged workers tend to have lower unemployment, probably because the middle-aged workers feel the responsibility of needing to have a job more heavily. Younger workers move in and out of jobs (and in and out of the labor force) more easily. Elderly workers have extremely low rates of unemployment, because those who do not have jobs often exit the labor force by retiring, and thus are not counted in the unemployment statistics. [\[link\]](#) (b) shows unemployment rates for women divided by age; the pattern for men is similar.

The unemployment rate for African-Americans is substantially higher than the rate for other racial or ethnic groups, a fact that surely reflects, to some extent, a pattern of discrimination that has constrained blacks’ labor market opportunities. However, the gaps between unemployment rates for whites and for blacks and Hispanics diminished in the 1990s, as shown in [\[link\]](#) (c). In fact, unemployment rates for blacks and Hispanics were at the lowest

levels for several decades in the mid-2000s before rising during the recent Great Recession.

Finally, those with less education typically suffer higher unemployment. In February 2015, for example, the unemployment rate for those with a college degree was 2.7%; for those with some college but not a four year degree, the unemployment rate was 5.1%; for high school graduates with no additional degree, the unemployment rate was 5.4%; and for those without a high school diploma, the unemployment rate was 8.4%. This pattern may arise because additional education offers better connections to the labor market and higher demand, or it may occur because the labor market opportunities for low-skilled workers are less attractive than the opportunities for the more highly-skilled. Because of lower pay, low-skilled workers may be less motivated to find jobs.

Key Concepts and Summary

The U.S. unemployment rate rises during periods of recession and depression, but falls back to the range of 4% to 6% when the economy is strong. The unemployment rate never falls to zero. Despite enormous growth in the size of the U.S. population and labor force in the twentieth century, along with other major trends like globalization and new technology, the unemployment rate shows no long-term rising trend.

Unemployment rates differ by group: higher for African-Americans and Hispanics than for whites; higher for less educated than more educated; higher for the young than the middle-aged. Women's unemployment rates used to be higher than men's, but in recent years men's and women's unemployment rates have been very similar. In recent years, unemployment rates in the United States have compared favorably with unemployment rates in most other high-income economies.

Self-Check Questions

Exercise:

Problem:

Over the long term, has the U.S. unemployment rate generally trended up, trended down, or remained at basically the same level?

Solution:

Over the long term, the U.S. unemployment rate has remained basically the same level.

Exercise:**Problem:**

Whose unemployment rates are commonly higher in the U.S. economy:

- a. Whites or nonwhites?
 - b. The young or the middle-aged?
 - c. College graduates or high school graduates?
-

Solution:

- a. Nonwhites
- b. The young
- c. High school graduates

Review Questions**Exercise:****Problem:**

Are U.S. unemployment rates typically higher, lower, or about the same as unemployment rates in other high-income countries?

Exercise:

Problem:

Are U.S. unemployment rates distributed evenly across the population?

Critical Thinking Questions**Exercise:****Problem:**

Is the higher unemployment rates for minority workers necessarily an indication of discrimination? What could be some other reasons for the higher unemployment rate?

Exercise:**Problem:**

While unemployment is highly negatively correlated with the level of economic activity, in the real world it responds with a lag. In other words, firms do not immediately lay off workers in response to a sales decline. They wait a while before responding. Similarly, firms do not immediately hire workers when sales pick up. What do you think accounts for the lag in response time?

Exercise:**Problem:**

Why do you think that unemployment rates are lower for individuals with more education?

What Causes Changes in Unemployment over the Short Run

By the end of this section, you will be able to:

- Analyze cyclical unemployment
- Explain the relationship between sticky wages and employment using various economic arguments
- Apply supply and demand models to unemployment and wages

We have seen that unemployment varies across times and places. What causes changes in unemployment? There are different answers in the short run and in the long run. Let's look at the short run first.

Cyclical Unemployment

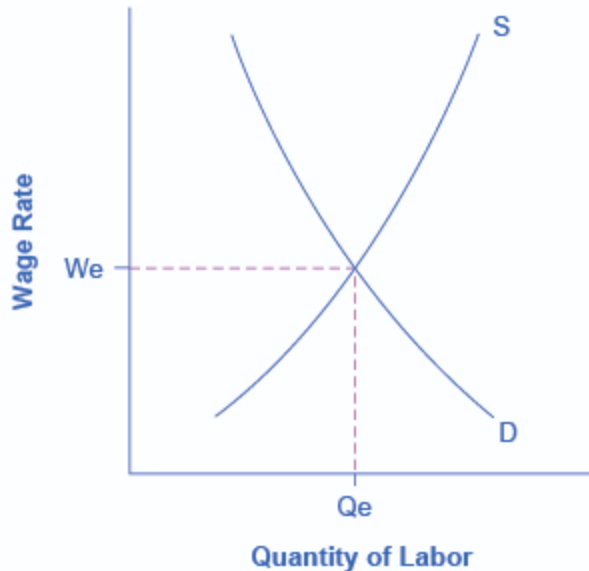
Let's make the plausible assumption that in the short run, from a few months to a few years, the quantity of hours that the average person is willing to work for a given wage does not change much, so the labor supply curve does not shift much. In addition, make the standard *ceteris paribus* assumption that there is no substantial short-term change in the age structure of the labor force, institutions and laws affecting the labor market, or other possibly relevant factors.

One primary determinant of the demand for labor from firms is how they perceive the state of the macro economy. If firms believe that business is expanding, then at any given wage they will desire to hire a greater quantity of labor, and the labor demand curve shifts to the right. Conversely, if firms perceive that the economy is slowing down or entering a recession, then they will wish to hire a lower quantity of labor at any given wage, and the labor demand curve will shift to the left. The variation in unemployment caused by the economy moving from expansion to recession or from recession to expansion (i.e. the business cycle) is known as **cyclical unemployment**.

From the standpoint of the supply-and-demand model of competitive and flexible labor markets, unemployment represents something of a puzzle. In a supply-and-demand model of a labor market, as illustrated in [\[link\]](#), the

labor market should move toward an equilibrium wage and quantity. At the equilibrium wage (W_e), the equilibrium quantity (Q_e) of labor supplied by workers should be equal to the quantity of labor demanded by employers.

The Unemployment and Equilibrium in the Labor Market



In a labor market with flexible wages, the equilibrium will occur at wage W_e and quantity Q_e , where the number of people looking for jobs (shown by S) equals the number of jobs available (shown by D).

One possibility for unemployment is that people who are unemployed are those who are not willing to work at the current equilibrium wage, say \$10 an hour, but would be willing to work at a higher wage, like \$20 per hour. The monthly Current Population Survey would count these people as unemployed, because they say they are ready and looking for work (at \$20 per hour). But from an economist's point of view, these people are choosing to be unemployed.

Probably a few people are unemployed because of unrealistic expectations about wages, but they do not represent the majority of the unemployed. Instead, unemployed people often have friends or acquaintances of similar

skill levels who are employed, and the unemployed would be willing to work at the jobs and wages similar to what is being received by those people. But the employers of their friends and acquaintances do not seem to be hiring. In other words, these people are involuntarily unemployed. What causes involuntary unemployment?

Why Wages Might Be Sticky Downward

If a labor market model with flexible wages does not describe unemployment very well—because it predicts that anyone willing to work at the going wage can always find a job—then it may prove useful to consider economic models in which wages are not flexible or adjust only very slowly. In particular, even though wage increases may occur with relative ease, wage decreases are few and far between.

One set of reasons why wages may be “sticky downward,” as economists put it, involves economic laws and institutions. For low-skilled workers being paid the minimum wage, it is illegal to reduce their wages. For union workers operating under a multiyear contract with a company, wage cuts might violate the contract and create a labor dispute or a strike. However, minimum wages and union contracts are not a sufficient reason why wages would be sticky downward for the U.S. economy as a whole. After all, out of the 150 million or so workers in the U.S. economy, only about 1.4 million—less than 2% of the total—are paid the minimum wage. Similarly, only about 12% of American wage and salary workers are represented by a labor union. In other high-income countries, more workers may have their wages determined by unions or the minimum wage may be set at a level that applies to a larger share of workers. But for the United States, these two factors combined affect only about one-fifth or less of the labor force.

Economists looking for reasons why wages might be sticky downwards have focused on factors that may characterize most labor relationships in the economy, not just a few. A number of different theories have been proposed, but they share a common tone.

These theories of why wages tend not to move downward differ in their logic and their implications, and figuring out the strengths and weaknesses

of each theory is an ongoing subject of research and controversy among economists. All tend to imply that wages will decline only very slowly, if at all, even when the economy or a business is having tough times. When wages are inflexible and unlikely to fall, then either short-run or long-run unemployment can result.

This helps to explain the connection noted earlier: that unemployment tends to rise in recessions and to decline during expansions. The overall state of the economy shifts the labor demand curve and, combined with wages that are sticky downwards, unemployment changes. The rise in unemployment that occurs because of a recession is cyclical unemployment.

Note: The St. Louis Federal Reserve Bank is the best resource for macroeconomic time series data, known as the Federal Reserve Economic Data (FRED). FRED provides complete data sets on various measures of the unemployment rate as well as the monthly Bureau of Labor Statistics report on the results of the household and employment surveys.

Key Concepts and Summary

Cyclical unemployment rises and falls with the business cycle. In a labor market with flexible wages, wages will adjust in such a market so that quantity demanded of labor always equals the quantity supplied of labor at the equilibrium wage. Many theories have been proposed for why wages might not be flexible, but instead may adjust only in a “sticky” way, especially when it comes to downward adjustments: implicit contracts, efficiency wage theory, adverse selection of wage cuts, insider-outsider model, and relative wage coordination.

Self-Check Questions

Exercise:

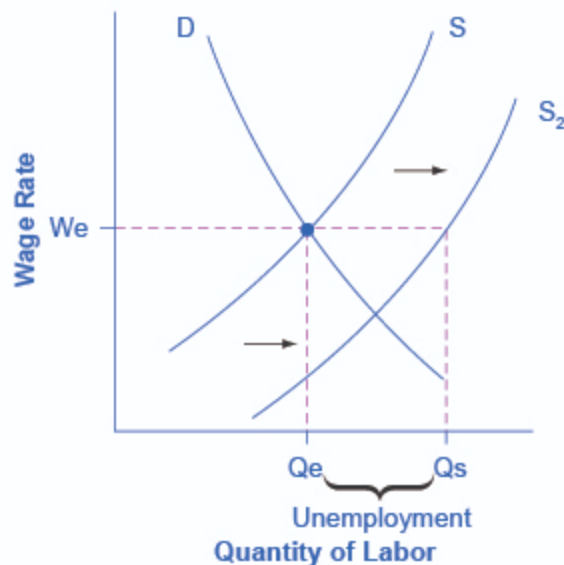
Problem:

Beginning in the 1970s and continuing for three decades, women entered the U.S. labor force in a big way. If we assume that wages are sticky in a downward direction, but that around 1970 the demand for labor equaled the supply of labor at the current wage rate, what do you imagine happened to the wage rate, employment, and unemployment as a result of increased labor force participation?

Solution:

Because of the influx of women into the labor market, the supply of labor shifts to the right. Since wages are sticky downward, the increased supply of labor causes an increase in people looking for jobs (Q_s), but no change in the number of jobs available (Q_e). As a result, unemployment increases by the amount of the increase in the labor supply. This can be seen in the following figure.

Over time, as labor demand grows, the unemployment will decline and eventually wages will begin to increase again. But this increase in labor demand goes beyond the scope of this problem.



Review Questions

Exercise:

Problem:

When would you expect cyclical unemployment to be rising? Falling?

Exercise:

Problem:

Why is there unemployment in a labor market with flexible wages?

Exercise:

Problem:

Name and explain some of the reasons why wages are likely to be sticky, especially in downward adjustments.

Critical Thinking Questions

Exercise:

Problem:

Do you think it is rational for workers to prefer sticky wages to wage cuts, when the consequence of sticky wages is unemployment for some workers? Why or why not? How do the reasons for sticky wages explained in this section apply to your argument?

Problems

Exercise:

Problem:

A government passes a family-friendly law that no companies can have evening, nighttime, or weekend hours, so that everyone can be home with their families during these times. Analyze the effect of this law using a demand and supply diagram for the labor market: first assuming that wages are flexible, and then assuming that wages are sticky downward.

Glossary

adverse selection of wage cuts argument

if employers reduce wages for all workers, the best will leave

cyclical unemployment

unemployment closely tied to the business cycle, like higher unemployment during a recession

efficiency wage theory

the theory that the productivity of workers, either individually or as a group, will increase if they are paid more

implicit contract

an unwritten agreement in the labor market that the employer will try to keep wages from falling when the economy is weak or the business is having trouble, and the employee will not expect huge salary increases when the economy or the business is strong

insider-outsider model

those already working for the firm are “insiders” who know the procedures; the other workers are “outsiders” who are recent or prospective hires

relative wage coordination argument

across-the-board wage cuts are hard for an economy to implement, and workers fight against them

What Causes Changes in Unemployment over the Long Run

By the end of this section, you will be able to:

- Explain frictional and structural unemployment
- Assess relationships between the natural rate of employment and potential real GDP, productivity, and public policy
- Identify recent patterns in the natural rate of employment
- Propose ways to combat unemployment

Cyclical unemployment explains why unemployment rises during a recession and falls during an economic expansion. But what explains the remaining level of unemployment even in good economic times? Why is the unemployment rate never zero? Even when the U.S. economy is growing strongly, the unemployment rate only rarely dips as low as 4%. Moreover, the discussion earlier in this chapter pointed out that unemployment rates in many European countries like Italy, France, and Germany have often been remarkably high at various times in the last few decades. Why does some level of unemployment persist even when economies are growing strongly? Why are unemployment rates continually higher in certain economies, through good economic years and bad? Economists have a term to describe the remaining level of unemployment that occurs even when the economy is healthy: it is called the **natural rate of unemployment**.

The Long Run: The Natural Rate of Unemployment

The natural rate of unemployment is not “natural” in the sense that water freezes at 32 degrees Fahrenheit or boils at 212 degrees Fahrenheit. It is not a physical and unchanging law of nature. Instead, it is only the “natural” rate because it is the unemployment rate that would result from the combination of economic, social, and political factors that exist at a time—assuming the economy was neither booming nor in recession. These forces include the usual pattern of companies expanding and contracting their workforces in a dynamic economy, social and economic forces that affect the labor market, or public policies that affect either the eagerness of people

to work or the willingness of businesses to hire. Let's discuss these factors in more detail.

Frictional Unemployment

In a market economy, some companies are always going broke for a variety of reasons: old technology; poor management; good management that happened to make bad decisions; shifts in tastes of consumers so that less of the firm's product is desired; a large customer who went broke; or tough domestic or foreign competitors. Conversely, other companies will be doing very well for just the opposite reasons and looking to hire more employees. In a perfect world, all of those who lost jobs would immediately find new ones. But in the real world, even if the number of job seekers is equal to the number of job vacancies, it takes time to find out about new jobs, to interview and figure out if the new job is a good match, or perhaps to sell a house and buy another in proximity to a new job. The unemployment that occurs in the meantime, as workers move between jobs, is called **frictional unemployment**. Frictional unemployment is not inherently a bad thing. It takes time on part of both the employer and the individual to match those looking for employment with the correct job openings. For individuals and companies to be successful and productive, you want people to find the job for which they are best suited, not just the first job offered.

In the mid-2000s, before the recession of 2008–2009, it was true that about 7% of U.S. workers saw their jobs disappear in any three-month period. But in periods of economic growth, these destroyed jobs are counterbalanced for the economy as a whole by a larger number of jobs created. In 2005, for example, there were typically about 7.5 million unemployed people at any given time in the U.S. economy. Even though about two-thirds of those unemployed people found a job in 14 weeks or fewer, the unemployment rate did not change much during the year, because those who found new jobs were largely offset by others who lost jobs.

The level of frictional unemployment will depend on how easy it is for workers to learn about alternative jobs, which may reflect the ease of communications about job prospects in the economy. The extent of frictional unemployment will also depend to some extent on how willing

people are to move to new areas to find jobs—which in turn may depend on history and culture.

Structural Unemployment

Another factor that influences the natural rate of unemployment is the amount of **structural unemployment**. The structurally unemployed are individuals who have no jobs because they lack skills valued by the labor market, either because demand has shifted away from the skills they do have, or because they never learned any skills. An example of the former would be the unemployment among aerospace engineers after the U.S. space program downsized in the 1970s. An example of the latter would be high school dropouts.

Some people worry that technology causes structural unemployment. In the past, new technologies have put lower skilled employees out of work, but at the same time they create demand for higher skilled workers to use the new technologies. Education seems to be the key in minimizing the amount of structural unemployment. Individuals who have degrees can be retrained if they become structurally unemployed. For people with no skills and little education, that option is more limited.

Natural Unemployment and Potential Real GDP

The natural unemployment rate is related to two other important concepts: full employment and potential real GDP. The economy is considered to be at full employment when the actual unemployment rate is equal to the natural unemployment. When the economy is at full employment, real GDP is equal to potential real GDP. By contrast, when the economy is below full employment, the unemployment rate is greater than the natural unemployment rate and real GDP is less than potential. Finally, when the economy is above full employment, then the unemployment rate is less than the natural unemployment rate and real GDP is greater than potential. Operating above potential is only possible for a short while, since it is analogous to all workers working overtime.

Productivity Shifts and the Natural Rate of Unemployment

Unexpected shifts in productivity can have a powerful effect on the natural rate of unemployment. Over time, the level of wages in an economy will be determined by the productivity of workers. After all, if a business paid workers more than could be justified by their productivity, the business will ultimately lose money and go bankrupt. Conversely, if a business tries to pay workers less than their productivity then, in a competitive labor market, other businesses will find it worthwhile to hire away those workers and pay them more.

However, adjustments of wages to productivity levels will not happen quickly or smoothly. Wages are typically reviewed only once or twice a year. In many modern jobs, it is difficult to measure productivity at the individual level. For example, how precisely would one measure the quantity produced by an accountant who is one of many people working in the tax department of a large corporation? Because productivity is difficult to observe, wage increases are often determined based on recent experience with productivity; if productivity has been rising at, say, 2% per year, then wages rise at that level as well. However, when productivity changes unexpectedly, it can affect the natural rate of unemployment for a time.

Average levels of unemployment will tend to be somewhat higher on average when productivity is unexpectedly low, and conversely, will tend to be somewhat lower on average when productivity is unexpectedly high. But over time, wages do eventually adjust to reflect productivity levels.

Public Policy and the Natural Rate of Unemployment

Public policy can also have a powerful effect on the natural rate of unemployment. On the supply side of the labor market, public policies to assist the unemployed can affect how eager people are to find work. For example, if a worker who loses a job is guaranteed a generous package of unemployment insurance, welfare benefits, food stamps, and government medical benefits, then the opportunity cost of being unemployed is lower and that worker will be less eager to seek a new job.

What seems to matter most is not just the amount of these benefits, but how long they last. A society that provides generous help for the unemployed that cuts off after, say, six months, may provide less of an incentive for unemployment than a society that provides less generous help that lasts for several years. Conversely, government assistance for job search or retraining can in some cases encourage people back to work sooner.

On the demand side of the labor market, government rules social institutions, and the presence of unions can affect the willingness of firms to hire. For example, if a government makes it hard for businesses to start up or to expand, by wrapping new businesses in bureaucratic red tape, then businesses will become more discouraged about hiring. Government regulations can make it harder to start a business by requiring that a new business obtain many permits and pay many fees, or by restricting the types and quality of products that can be sold. Other government regulations, like zoning laws, may limit where business can be done, or whether businesses are allowed to be open during evenings or on Sunday.

Whatever defenses may be offered for such laws in terms of social value—like the value some Christians place on not working on Sunday—these kinds of restrictions impose a barrier between some willing workers and other willing employers, and thus contribute to a higher natural rate of unemployment. Similarly, if government makes it difficult to fire or lay off workers, businesses may react by trying not to hire more workers than strictly necessary—since laying these workers off would be costly and difficult. High minimum wages may discourage businesses from hiring low-skill workers. Government rules may encourage and support powerful unions, which can then push up wages for union workers, but at a cost of discouraging businesses from hiring those workers.

Key Concepts and Summary

The natural rate of unemployment is the rate of unemployment that would be caused by the economic, social, and political forces in the economy even when the economy is not in a recession. These factors include the frictional unemployment that occurs when people are put out of work for a time by the shifts of a dynamic and changing economy and any laws concerning

conditions of hiring and firing have the undesired side effect of discouraging job formation. They also include structural unemployment, which occurs when demand shifts permanently away from a certain type of job skill.

Self-Check Questions

Exercise:

Problem:

Is the increase in labor force participation rates among women better thought of as causing an increase in cyclical unemployment or an increase in the natural rate of unemployment? Why?

Solution:

The increase in labor supply was a social demographic trend—it was not caused by the economy falling into a recession. Therefore, the influx of women into the work force increased the natural rate of unemployment.

Exercise:

Problem:

Many college students graduate from college before they have found a job. When graduates begin to look for a job, they are counted as what category of unemployed?

Solution:

New entrants to the labor force, whether from college or otherwise, are counted as frictionally unemployed until they find a job.

Review Questions

Exercise:

Problem:

What term describes the remaining level of unemployment that occurs even when the economy is healthy?

Exercise:

Problem:

What forces create the natural rate of unemployment for an economy?

Exercise:

Problem:

Would you expect the natural rate of unemployment to be roughly the same in different countries?

Exercise:

Problem:

Would you expect the natural rate of unemployment to remain the same within one country over the long run of several decades?

Exercise:

Problem:

What is frictional unemployment? Give examples of frictional unemployment.

Exercise:

Problem:

What is structural unemployment? Give examples of structural unemployment.

Exercise:

Problem:

After several years of economic growth, would you expect the unemployment in an economy to be mainly cyclical or mainly due to the natural rate of unemployment? Why?

Exercise:**Problem:**

What type of unemployment (cyclical, frictional, or structural) applies to each of the following:

- A. landscapers laid off in response to drop in new housing construction during a recession.
- B. coal miners laid off due to EPA regulations that shut down coal fired power
- C. a financial analyst who quits his/her job in Chicago and is pursuing similar work in Arizona
- D. printers laid off due to drop in demand for printed catalogues and flyers as firms go the internet to promote and advertise their products.
- E. factory workers in the U.S. laid off as the plants shut down and move to Mexico and Ireland.

Critical Thinking Questions**Exercise:****Problem:**

Under what condition would a decrease in unemployment be bad for the economy?

Exercise:

Problem:

Under what condition would an increase in the unemployment rate be a positive sign?

Exercise:

Problem:

As the baby boom generation retires, the ratio of retirees to workers will increase noticeably. How will this affect the Social Security program? How will this affect the standard of living of the average American?

Exercise:

Problem:

Unemployment rates have been higher in many European countries in recent decades than in the United States. Is the main reason for this long-term difference in unemployment rates more likely to be cyclical unemployment or the natural rate of unemployment? Explain briefly.

Exercise:

Problem:

Is it desirable to pursue a goal of zero unemployment? Why or why not?

Exercise:

Problem:

Is it desirable to eliminate natural unemployment? Why or why not?
Hint: Think about what our economy would look like today and what assumptions would have to be met to have a zero rate of natural unemployment.

Exercise:

Problem:

The U.S. unemployment rate increased from 4.6% in July 2001 to 5.9% by June 2002. Without studying the subject in any detail, would you expect that a change of this kind is more likely to be due to cyclical unemployment or a change in the natural rate of unemployment? Why?

Problems**Exercise:****Problem:**

As the baby boomer generation retires, what should happen to wages and employment? Can you show this graphically?

References

Bureau of Labor Statistics. Labor Force Statistics from the Current Population Survey. *Accessed March 6, 2015*
<http://data.bls.gov/timeseries/LNS14000000>.

Cappelli, P. (20 June 2012). "Why Good People Can't Get Jobs: Chasing After the Purple Squirrel."
<http://knowledge.wharton.upenn.edu/article.cfm?articleid=3027>.

Glossary

frictional unemployment

unemployment that occurs as workers move between jobs

natural rate of unemployment

the unemployment rate that would exist in a growing and healthy economy from the combination of economic, social, and political factors that exist at a given time

structural unemployment

unemployment that occurs because individuals lack skills valued by employers

Introduction to Inflation
class="introduction"
Big Bucks in Zimbabwe

This bill was worth
100 billion
Zimbabwean dollars
when issued in 2008.
There were even bills
issued with a face
value of 100 trillion
Zimbabwean dollars.
The bills had
\$100,000,000,000,000
written on them.
Unfortunately, they
were almost
worthless. At one
point, 621,984,228
Zimbabwean dollars
were equal to one
U.S. dollar.
Eventually, the
country abandoned its
own currency and
allowed foreign
currency to be used
for purchases. (Credit:
modification of work
by Samantha
Marx/Flickr Creative
Commons)



Note:

A \$550 Million Loaf of Bread?

If you were born within the last three decades in the United States, Canada, or many other countries in the developed world, you probably have no real experience with a high rate of inflation. Inflation is when most prices in an entire economy are rising. But there is an extreme form of inflation called hyperinflation. This occurred in Germany between 1921 and 1928, and more recently in Zimbabwe between 2008 and 2009. In November of 2008, Zimbabwe had an inflation rate of 79.6 billion percent. In contrast, in 2014, the United States had an average annual rate of inflation of 1.6%. Zimbabwe's inflation rate was so high it is difficult to comprehend. So, let's put it into context. It is equivalent to price increases of 98% per day. This means that, from one day to the next, prices essentially double. What is life like in an economy afflicted with hyperinflation? Not like anything you are familiar with. Prices for commodities in Zimbabwean dollars were adjusted several times *each day*. There was no desire to hold on to currency since it lost value by the minute. The people there spent a great deal of time getting rid of any cash they acquired by purchasing whatever food or other commodities they could find. At one point, a loaf of bread cost 550 million Zimbabwean dollars. Teachers were paid in the trillions a month; however this was equivalent to only one U.S. dollar a day. At its

height, it took 621,984,228 Zimbabwean dollars to purchase one U.S. dollar.

Government agencies had no money to pay their workers so they started printing money to pay their bills rather than raising taxes. Rising prices caused the government to enact price controls on private businesses, which led to shortages and the emergence of black markets. In 2009, the country abandoned its currency and allowed foreign currencies to be used for purchases.

How does this happen? How can both government and the economy fail to function at the most basic level? Before we consider these extreme cases of hyperinflation, let's first look at inflation itself.

Note:

Introduction to Inflation

In this chapter, you will learn about:

- Tracking Inflation
- How Changes in the Cost of Living are Measured
- How the U.S. and Other Countries Experience Inflation
- The Confusion Over Inflation
- Indexing and Its Limitations

Inflation is a general and ongoing rise in the level of prices in an entire economy. Inflation does not refer to a change in relative prices. A relative price change occurs when you see that the price of tuition has risen, but the price of laptops has fallen. Inflation, on the other hand, means that there is pressure for prices to rise in most markets in the economy. In addition, price increases in the supply-and-demand model were one-time events, representing a shift from a previous equilibrium to a new one. Inflation implies an ongoing rise in prices. If inflation happened for one year and then stopped—well, then it would not be inflation any more.

This chapter begins by showing how to combine prices of individual goods and services to create a measure of overall inflation. It discusses the historical and recent experience of inflation, both in the United States and in other countries around the world. Other chapters have sometimes included a note under an exhibit or a parenthetical reminder in the text saying that the numbers have been adjusted for inflation. In this chapter, it is time to show how to use inflation statistics to adjust other economic variables, so that you can tell how much of, say, the rise in GDP over different periods of time can be attributed to an actual increase in the production of goods and services and how much should be attributed to the fact that prices for most things have risen.

Inflation has consequences for people and firms throughout the economy, in their roles as lenders and borrowers, wage-earners, taxpayers, and consumers. The chapter concludes with a discussion of some imperfections and biases in the inflation statistics, and a preview of policies for fighting inflation that will be discussed in other chapters.

Tracking Inflation

By the end of this section, you will be able to:

- Calculate the annual rate of inflation
- Explain and use index numbers and base years when simplifying the total quantity spent over a year for products
- Calculate inflation rates using index numbers

Dinner table conversations where you might have heard about inflation usually entail reminiscing about when “everything seemed to cost so much less. You used to be able to buy three gallons of gasoline for a dollar and then go see an afternoon movie for another dollar.” [\[link\]](#) compares some prices of common goods in 1970 and 2014. Of course, the average prices shown in this table may not reflect the prices where you live. The cost of living in New York City is much higher than in Houston, Texas, for example. In addition, certain products have evolved over recent decades. A new car in 2014, loaded with antipollution equipment, safety gear, computerized engine controls, and many other technological advances, is a more advanced machine (and more fuel efficient) than your typical 1970s car. However, put details like these to one side for the moment, and look at the overall pattern. The primary reason behind the price rises in [\[link\]](#)—and all the price increases for the other products in the economy—is not specific to the market for housing or cars or gasoline or movie tickets. Instead, it is part of a general rise in the level of all prices. In 2014, \$1 had about the same purchasing power in overall terms of goods and services as 18 cents did in 1972, because of the amount of inflation that has occurred over that time period.

Items	1970	2014
Pound of ground beef	\$0.66	\$4.16
Pound of butter	\$0.87	\$2.93
Movie ticket	\$1.55	\$8.17
Sales price of new home (median)	\$22,000	\$280,000
New car	\$3,000	\$32,531
Gallon of gasoline	\$0.36	\$3.36
Average hourly wage for a manufacturing worker	\$3.23	\$19.55
Per capita GDP	\$5,069	\$53,041.98

Price Comparisons, 1970 and 2014(Sources: See chapter References at end of book.)

Moreover, the power of inflation does not affect just goods and services, but wages and income levels, too. The second-to-last row of [\[link\]](#) shows that the average hourly wage for a manufacturing worker increased nearly six-fold from 1970 to 2014. Sure, the average worker in 2014 is better educated and more productive than the average worker in 1970—but not six times more productive. Sure, per capita GDP increased substantially from 1970 to 2014, but is the average person in the U.S. economy really more than eight times better off in just 44 years? Not likely.

A modern economy has millions of goods and services whose prices are continually quivering in the breezes of supply and demand. How can all of these shifts in price be boiled down to a single inflation rate? As with many problems in economic measurement, the conceptual answer is reasonably straightforward: Prices of a variety of

goods and services are combined into a single price level; the inflation rate is simply the percentage change in the price level. Applying the concept, however, involves some practical difficulties.

The Price of a Basket of Goods

To calculate the price level, economists begin with the concept of a **basket of goods and services**, consisting of the different items individuals, businesses, or organizations typically buy. The next step is to look at how the prices of those items change over time. In thinking about how to combine individual prices into an overall price level, many people find that their first impulse is to calculate the average of the prices. Such a calculation, however, could easily be misleading because some products matter more than others.

Changes in the prices of goods for which people spend a larger share of their incomes will matter more than changes in the prices of goods for which people spend a smaller share of their incomes. For example, an increase of 10% in the rental rate on housing matters more to most people than whether the price of carrots rises by 10%. To construct an overall measure of the price level, economists compute a weighted average of the prices of the items in the basket, where the weights are based on the actual quantities of goods and services people buy. The following Work It Out feature walks you through the steps of calculating the annual rate of inflation based on a few products.

Note:
Calculating an Annual Rate of Inflation
Consider the simple basket of goods with only three items, represented in [\[link\]](#). Say that in any given month, a college student spends money on 20 hamburgers, one bottle of aspirin, and five movies. Prices for these items over four years are given in the table through each time period (Pd). Prices of some goods in the basket may rise while others fall. In this example, the price of aspirin does not change over the four years, while movies increase in price and hamburgers bounce up and down. Each year, the cost of buying the given basket of goods at the prices prevailing at that time is shown.

Items	Hamburger	Aspirin	Movies	Total	Inflation Rate
Qty	20	1 bottle	5	-	-
(Pd 1) Price	\$3.00	\$10.00	\$6.00	-	-
(Pd 1) Amount Spent	\$60.00	\$10.00	\$30.00	\$100.00	-
(Pd 2) Price	\$3.20	\$10.00	\$6.50	-	-
(Pd 2) Amount Spent	\$64.00	\$10.00	\$32.50	\$106.50	6.5%
(Pd 3) Price	\$3.10	\$10.00	\$7.00	-	-
(Pd 3) Amount Spent	\$62.00	\$10.00	\$35.00	\$107.00	0.5%
(Pd 4) Price	\$3.50	\$10.00	\$7.50	-	-
(Pd 4) Amount Spent	\$70.00	\$10.00	\$37.50	\$117.50	9.8%

A College Student’s Basket of Goods

To calculate the annual rate of inflation in this example:

Step 1. Find the percentage change in the cost of purchasing the overall basket of goods between the time periods. The general equation for percentage changes between two years, whether in the context of inflation or in any other calculation, is:

Equation:

$$\frac{(\text{Level in new year} - \text{Level in previous year})}{\text{Level in previous year}} = \text{Percentage change}$$

Step 2. From period 1 to period 2, the total cost of purchasing the basket of goods in [\[link\]](#) rises from \$100 to \$106.50. Therefore, the percentage change over this time—the inflation rate—is:

Equation:

$$\frac{(106.50 - 100)}{100.0} = 0.065 = 6.5\%$$

Step 3. From period 2 to period 3, the overall change in the cost of purchasing the basket rises from \$106.50 to \$107. Thus, the inflation rate over this time, again calculated by the percentage change, is approximately:

Equation:

$$\frac{(107 - 106.50)}{106.50} = 0.0047 = 0.47\%$$

Step 4. From period 3 to period 4, the overall cost rises from \$107 to \$117.50. The inflation rate is thus:

Equation:

$$\frac{(117.50 - 107)}{107} = 0.098 = 9.8\%$$

This calculation of the change in the total cost of purchasing a basket of goods takes into account how much is spent on each good. Hamburgers are the lowest-priced good in this example, and aspirin is the highest-priced. If an individual buys a greater quantity of a low-price good, then it makes sense that changes in the price of that good should have a larger impact on the buying power of that person's money. The larger impact of hamburgers shows up in the "amount spent" row, where, in all time periods, hamburgers are the largest item within the amount spent row.

Index Numbers

The numerical results of a calculation based on a basket of goods can get a little messy. The simplified example in [\[link\]](#) has only three goods and the prices are in even dollars, not numbers like 79 cents or \$124.99. If the list of products was much longer, and more realistic prices were used, the total quantity spent over a year might be some messy-looking number like \$17,147.51 or \$27,654.92.

To simplify the task of interpreting the price levels for more realistic and complex baskets of goods, the price level in each period is typically reported as an **index number**, rather than as the dollar amount for buying the basket of goods. Price indices are created to calculate an overall average change in relative prices over time. To convert the money spent on the basket to an index number, economists arbitrarily choose one year to be the **base year**, or starting point from which we measure changes in prices. The base year, by definition, has an index number equal to 100. This sounds complicated, but it is really a simple math trick. In the example above, say that time period 3 is chosen as the base year. Since the total amount of spending in that year is \$107, we divide that amount by itself (\$107) and multiply by 100. Mathematically, that is equivalent to dividing \$107 by 100, or \$1.07. Doing either will give us an index in the base year of 100. Again, this is because the index number in the base year *always* has to have a value of 100. Then, to figure out the values of the index number for the other years, we divide the dollar amounts for the other years by 1.07 as well. Note also that the dollar signs cancel out so that index numbers have no units.

Calculations for the other values of the index number, based on the example presented in [\[link\]](#) are shown in [\[link\]](#). Because the index numbers are calculated so that they are in exactly the same proportion as the total dollar cost of

purchasing the basket of goods, the inflation rate can be calculated based on the index numbers, using the percentage change formula. So, the inflation rate from period 1 to period 2 would be

Equation:

$$\frac{(99.5 - 93.4)}{93.4} = 0.065 = 6.5\%$$

This is the same answer that was derived when measuring inflation based on the dollar cost of the basket of goods for the same time period.

	Total Spending	Index Number	Inflation Rate Since Previous Period
Period 1	\$100	$\frac{100}{1.07} = 93.4$	
Period 2	\$106.50	$\frac{106.50}{1.07} = 99.5$	$\frac{(99.5 - 93.4)}{93.4} = 0.065 = 6.5\%$
Period 3	\$107	$\frac{107}{1.07} = 100.0$	$\frac{100 - 99.5}{99.5} = 0.005 = 0.5\%$
Period 4	\$117.50	$\frac{117.50}{1.07} = 109.8$	$\frac{109.8 - 100}{100} = 0.098 = 9.8\%$

Calculating Index Numbers When Period 3 is the Base Year

If the inflation rate is the same whether it is based on dollar values or index numbers, then why bother with the index numbers? The advantage is that indexing allows easier eyeballing of the inflation numbers. If you glance at two index numbers like 107 and 110, you know automatically that the rate of inflation between the two years is about, but not quite exactly equal to, 3%. By contrast, imagine that the price levels were expressed in absolute dollars of a large basket of goods, so that when you looked at the data, the numbers were \$19,493.62 and \$20,009.32. Most people find it difficult to eyeball those kinds of numbers and say that it is a change of about 3%. However, the two numbers expressed in absolute dollars are exactly in the same proportion of 107 to 110 as the previous example. If you're wondering why simple subtraction of the index numbers wouldn't work, read the following Clear It Up feature.

Note:

Why do you not just subtract index numbers?

A word of warning: When a price index moves from, say, 107 to 110, the rate of inflation is not *exactly* 3%. Remember, the inflation rate is not derived by subtracting the index numbers, but rather through the percentage-change calculation. The precise inflation rate as the price index moves from 107 to 110 is calculated as $(110 - 107) / 107 = 0.028 = 2.8\%$. When the base year is fairly close to 100, a quick subtraction is not a terrible shortcut to calculating the inflation rate—but when precision matters down to tenths of a percent, subtracting will not give the right answer.

Two final points about index numbers are worth remembering. First, index numbers have no dollar signs or other units attached to them. Although index numbers can be used to calculate a percentage inflation rate, the index numbers themselves do not have percentage signs. Index numbers just mirror the proportions found in other data. They transform the other data so that the data are easier to work with.

Second, the choice of a base year for the index number—that is, the year that is automatically set equal to 100—is arbitrary. It is chosen as a starting point from which changes in prices are tracked. In the official inflation statistics, it is common to use one base year for a few years, and then to update it, so that the base year of 100 is relatively close to the present. But any base year that is chosen for the index numbers will result in exactly the same inflation rate. To see this in the previous example, imagine that period 1, when total spending was \$100, was also chosen as the base year, and given an index number of 100. At a glance, you can see that the index numbers would now exactly match the dollar figures, the inflation rate in the first period would be 6.5%, and so on.

Now that we see how indexes work to track inflation, the next module will show us how the cost of living is measured.

Note: Watch this [video](#) from the cartoon *Duck Tales* to view a mini-lesson on inflation.

Key Concepts and Summary

The price level is measured by using a basket of goods and services and calculating how the total cost of buying that basket of goods will increase over time. The price level is often expressed in terms of index numbers, which transform the cost of buying the basket of goods and services into a series of numbers in the same proportion to each other, but with an arbitrary base year of 100. The rate of inflation is measured as the percentage change between price levels or index numbers over time.

Self-Check Questions

Exercise:

Problem:

[\[link\]](#) shows the prices of fruit purchased by the typical college student from 2001 to 2004. What is the amount spent each year on the “basket” of fruit with the quantities shown in column 2?

Items	Qty	(2001) Price	(2001) Amount Spent	(2002) Price	(2002) Amount Spent	(2003) Price	(2003) Amount Spent	(2004) Price
Apples	10	\$0.50		\$0.75		\$0.85		\$0.85
Bananas	12	\$0.20		\$0.25		\$0.25		\$0.25
Grapes	2	\$0.65		\$0.70		\$0.90		\$0.90
Raspberries	1	\$2.00		\$1.90		\$2.05		\$2.10
Total								

Solution:

To compute the amount spent on each fruit in each year, you multiply the quantity of each fruit by the price.

- 10 apples \times 50 cents each = \$5.00 spent on apples in 2001.
- 12 bananas \times 20 cents each = \$2.40 spent on bananas in 2001.
- 2 bunches of grapes at 65 cents each = \$1.30 spent on grapes in 2001.
- 1 pint of raspberries at \$2 each = \$2.00 spent on raspberries in 2001.

Adding up the amounts gives you the total cost of the fruit basket. The total cost of the fruit basket in 2001 was $\$5.00 + \$2.40 + \$1.30 + \$2.00 = \$10.70$. The total costs for all the years are shown in the following table.

2001	2002	2003	2004
\$10.70	\$13.80	\$15.35	\$16.31

Exercise:

Problem: Construct the price index for a “fruit basket” in each year using 2003 as the base year.

Solution:

If 2003 is the base year, then the index number has a value of 100 in 2003. To transform the cost of a fruit basket each year, we divide each year’s value by \$15.35, the value of the base year, and then multiply the result by 100. The price index is shown in the following table.

2001	2002	2003	2004
69.71	89.90	100.00	106.3

Note that the base year has a value of 100; years before the base year have values less than 100; and years after have values more than 100.

Exercise:

Problem: Compute the inflation rate for fruit prices from 2001 to 2004.

Solution:

The inflation rate is calculated as the percentage change in the price index from year to year. For example, the inflation rate between 2001 and 2002 is $(84.61 - 69.71) / 69.71 = 0.2137 = 21.37\%$. The inflation rates for all the years are shown in the last row of the following table, which includes the two previous answers.

Items	Qty	(2001) Price	(2001) Amount Spent	(2002) Price	(2002) Amount Spent	(2003) Price	(2003) Amount Spent	(2004) Price
Apples	10	\$0.50	\$5.00	\$0.75	\$7.50	\$0.85	\$8.50	\$0.85
Bananas	12	\$0.20	\$2.40	\$0.25	\$3.00	\$0.25	\$3.00	\$0.25
Grapes	2	\$0.65	\$1.30	\$0.70	\$1.40	\$0.90	\$1.80	\$0.90
Raspberries	1	\$2.00	\$2.00	\$1.90	\$1.90	\$2.05	\$2.05	\$2.15
Total			\$10.70		\$13.80		\$15.35	
Price Index			69.71		84.61		100.00	
Inflation Rate					21.37%		18.19%	

Exercise:

Problem:

Edna is living in a retirement home where most of her needs are taken care of, but she has some discretionary spending. Based on the basket of goods in [\[link\]](#), by what percentage does Edna's cost of living increase between time 1 and time 2?

Items	Quantity	(Time 1) Price	(Time 2) Price
Gifts for grandchildren	12	\$50	\$60
Pizza delivery	24	\$15	\$16
Blouses	6	\$60	\$50
Vacation trips	2	\$400	\$420

Solution:

Begin by calculating the total cost of buying the basket in each time period, as shown in the following table.

Items	Quantity	(Time 1) Price	(Time 1) Total Cost	(Time 2) Price	(Time 2) Total Cost
-------	----------	-------------------	------------------------	-------------------	------------------------

Items	Quantity	(Time 1) Price	(Time 1) Total Cost	(Time 2) Price	(Time 2) Total Cost
Gifts	12	\$50	\$600	\$60	\$720
Pizza	24	\$15	\$360	\$16	\$384
Blouses	6	\$60	\$360	\$50	\$300
Trips	2	\$400	\$800	\$420	\$840
Total Cost			\$2,120		\$2,244

The rise in cost of living is calculated as the percentage increase:

$$(2244 - 2120) / 2120 = 0.0585 = 5.85\%.$$

Review Questions

Exercise:

Problem: How is a basket of goods and services used to measure the price level?

Exercise:

Problem: Why are index numbers used to measure the price level rather than dollar value of goods?

Exercise:

Problem: What is the difference between the price level and the rate of inflation?

Critical Thinking Question

Exercise:

Problem:

Inflation rates, like most statistics, are imperfect measures. Can you identify some ways that the inflation rate for fruit does not perfectly capture the rising price of fruit?

Problems

Exercise:

Problem:

The index number representing the price level changes from 110 to 115 in one year, and then from 115 to 120 the next year. Since the index number increases by five each year, is five the inflation rate each year? Is the inflation rate the same each year? Explain your answer.

Exercise:

Problem:

The total price of purchasing a basket of goods in the United Kingdom over four years is: year 1=£940, year 2=£970, year 3=£1000, and year 4=£1070. Calculate two price indices, one using year 1 as the base year (set equal to 100) and the other using year 4 as the base year (set equal to 100). Then, calculate the inflation rate based on the first price index. If you had used the other price index, would you get a different inflation rate? If you are unsure, do the calculation and find out.

References

Sources for [\[link\]](#):

http://www.eia.gov/dnav/pet/pet_pri_gnd_a_epmr_pte_dpgal_w.htm

<http://data.bls.gov/cgi-bin/surveymost?ap>

<http://www.bls.gov/ro3/apmw.htm>

<http://www.autoblog.com/2014/03/12/who-can-afford-the-average-car-price-only-folks-in-washington/>

<https://www.census.gov/construction/nrs/pdf/uspricemon.pdf>

<http://www.bls.gov/news.release/empsit.t24.htm>

<http://variety.com/2015/film/news/movie-ticket-prices-increased-in-2014-1201409670/>

US Inflation Calculator. "Historical Inflation Rates: 1914-2013." Accessed March 4, 2015.

<http://www.usinflationcalculator.com/inflation/historical-inflation-rates/>.

Glossary

base year

arbitrary year whose value as an index number is defined as 100; inflation from the base year to other years can easily be seen by comparing the index number in the other year to the index number in the base year—for example, 100; so, if the index number for a year is 105, then there has been exactly 5% inflation between that year and the base year

basket of goods and services

a hypothetical group of different items, with specified quantities of each one meant to represent a “typical” set of consumer purchases, used as a basis for calculating how the price level changes over time

index number

a unit-free number derived from the price level over a number of years, which makes computing inflation rates easier, since the index number has values around 100

inflation

a general and ongoing rise in the level of prices in an economy

How Changes in the Cost of Living are Measured

By the end of this section, you will be able to:

- Use the Consumer Price Index (CPI) to calculate U.S. inflation rates
- Identify several ways the Bureau of Labor Statistics avoids biases in the Consumer Price Index (CPI)
- Differentiate among the Consumer Price Index (CPI), the Producer Price Index (PPI), the International Price Index, the Employment Cost Index, and the GDP deflator.

The most commonly cited measure of inflation in the United States is the **Consumer Price Index (CPI)**. The CPI is calculated by government statisticians at the U.S. Bureau of Labor Statistics based on the prices in a fixed basket of goods and services that represents the purchases of the average family of four. In recent years, the statisticians have paid considerable attention to a subtle problem: that the change in the total cost of buying a fixed basket of goods and services over time is conceptually not quite the same as the change in the cost of living, because the cost of living represents how much it costs for a person to feel that his or her consumption provides an equal level of satisfaction or utility.

To understand the distinction, imagine that over the past 10 years, the cost of purchasing a fixed basket of goods increased by 25% and your salary also increased by 25%. Has your personal standard of living held constant? If you do not necessarily purchase an identical fixed basket of goods every year, then an inflation calculation based on the cost of a fixed basket of goods may be a misleading measure of how your cost of living has changed. Two problems arise here: substitution bias and quality/new goods bias.

When the price of a good rises, consumers tend to purchase less of it and to seek out substitutes instead. Conversely, as the price of a good falls, people will tend to purchase more of it. This pattern implies that goods with generally rising prices should tend over time to become less important in the overall basket of goods used to calculate inflation, while goods with falling prices should tend to become more important. Consider, as an example, a rise in the price of peaches by \$100 per pound. If consumers

were utterly inflexible in their demand for peaches, this would lead to a big rise in the price of food for consumers. Alternatively, imagine that people are utterly indifferent to whether they have peaches or other types of fruit. Now, if peach prices rise, people completely switch to other fruit choices and the average price of food does not change at all. A fixed and unchanging basket of goods assumes that consumers are locked into buying exactly the same goods, regardless of price changes—not a very likely assumption. Thus, **substitution bias**—the rise in the price of a fixed basket of goods over time—tends to overstate the rise in a consumer's true cost of living, because it does not take into account that the person can substitute away from goods whose relative prices have risen.

The other major problem in using a fixed basket of goods as the basis for calculating inflation is how to deal with the arrival of improved versions of older goods or altogether new goods. Consider the problem that arises if a cereal is improved by adding 12 essential vitamins and minerals—and also if a box of the cereal costs 5% more. It would clearly be misleading to count the entire resulting higher price as inflation, because the new price is being charged for a product of higher (or at least different) quality. Ideally, one would like to know how much of the higher price is due to the quality change, and how much of it is just a higher price. The Bureau of Labor Statistics, which is responsible for the computation of the Consumer Price Index, must deal with these difficulties in adjusting for quality changes.

A new product can be thought of as an extreme improvement in quality—from something that did not exist to something that does. However, the basket of goods that was fixed in the past obviously does not include new goods created since then. The basket of goods and services used in the Consumer Price Index (CPI) is revised and updated over time, and so new products are gradually included. But the process takes some time. For example, room air conditioners were widely sold in the early 1950s, but were not introduced into the basket of goods behind the Consumer Price Index until 1964. The VCR and personal computer were available in the late 1970s and widely sold by the early 1980s, but did not enter the CPI basket of goods until 1987. By 1996, there were more than 40 million cellular phone subscribers in the United States—but cell phones were not

yet part of the CPI basket of goods. The parade of inventions has continued, with the CPI inevitably lagging a few years behind.

The arrival of new goods creates problems with respect to the accuracy of measuring inflation. The reason people buy new goods, presumably, is that the new goods offer better value for money than existing goods. Thus, if the price index leaves out new goods, it overlooks one of the ways in which the cost of living is improving. In addition, the price of a new good is often higher when it is first introduced and then declines over time. If the new good is not included in the CPI for some years, until its price is already lower, the CPI may miss counting this price decline altogether. Taking these arguments together, the **quality/new goods bias** means that the rise in the price of a fixed basket of goods over time tends to overstate the rise in a consumer's true cost of living, because it does not take into account how improvements in the quality of existing goods or the invention of new goods improves the standard of living. The following Clear It Up feature is a must-read on how the CPI is comprised and calculated.

Note:

How do U.S. government statisticians measure the Consumer Price Index? When the U.S. Bureau of Labor Statistics (BLS) calculates the Consumer Price Index, the first task is to decide on a basket of goods that is representative of the purchases of the average household. This is done by using the Consumer Expenditure Survey, a national survey of about 7,000 households, which provides detailed information on spending habits. Consumer expenditures are broken up into eight major groups, shown below, which in turn are broken up into more than 200 individual item categories. The BLS currently uses 1982–1984 as the base period. For each of the 200 individual expenditure items, the BLS chooses several hundred very specific examples of that item and looks at the prices of those examples. So, in figuring out the “breakfast cereal” item under the overall category of “foods and beverages,” the BLS picks several hundred examples of breakfast cereal. One example might be the price of a 24-oz. box of a particular brand of cereal sold at a particular store. The specific products and sizes and stores chosen are statistically selected to reflect

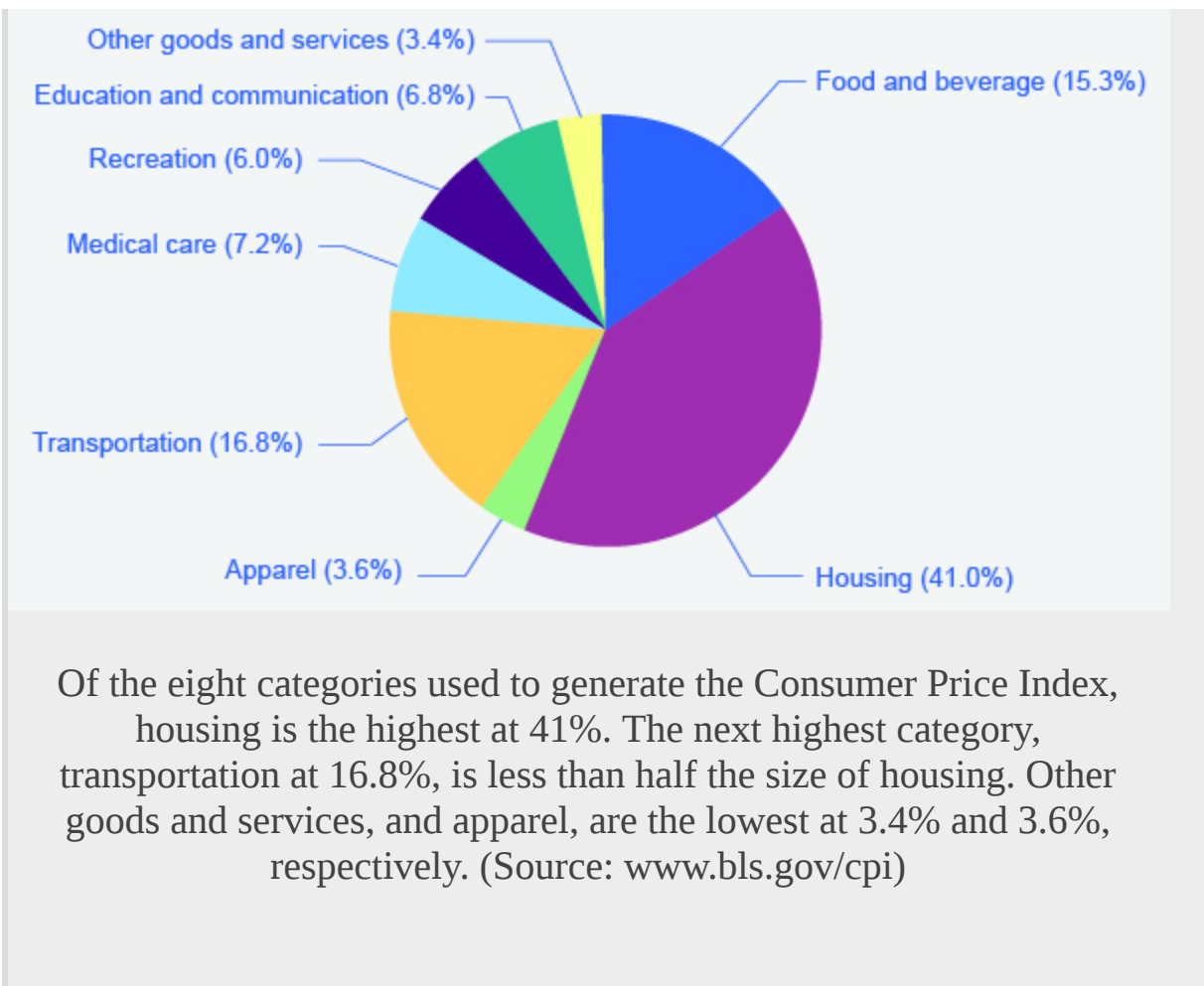
what people buy and where they shop. The basket of goods in the Consumer Price Index thus consists of about 80,000 products; that is, several hundred specific products in over 200 broad-item categories. About one-quarter of these 80,000 specific products are rotated out of the sample each year, and replaced with a different set of products.

The next step is to collect data on prices. Data collectors visit or call about 23,000 stores in 87 urban areas all over the United States every month to collect prices on these 80,000 specific products. A survey of 50,000 landlords or tenants is also carried out to collect information about rents. The Consumer Price Index is then calculated by taking the 80,000 prices of individual products and combining them, using weights (as shown in [\[link\]](#)) determined by the quantities of these products that people buy and allowing for factors like substitution between goods and quality improvements, into price indices for the 200 or so overall items. Then, the price indices for the 200 items are combined into an overall Consumer Price Index. According the Consumer Price Index website, there are eight categories used by data collectors:

The Eight Major Categories in the Consumer Price Index

1. Food and beverages (breakfast cereal, milk, coffee, chicken, wine, full-service meals, and snacks)
2. Housing (renter's cost of housing, homeowner's cost of housing, fuel oil, bedroom furniture)
3. Apparel (men's shirts and sweaters, women's dresses, jewelry)
4. Transportation (new vehicles, airline fares, gasoline, motor vehicle insurance)
5. Medical care (prescription drugs and medical supplies, physicians' services, eyeglasses and eye care, hospital services)
6. Recreation (televisions, cable television, pets and pet products, sports equipment, admissions)
7. Education and communication (college tuition, postage, telephone services, computer software and accessories)
8. Other goods and services (tobacco and smoking products, haircuts and other personal services, funeral expenses)

The Weighting of CPI Components



Practical Solutions for the Substitution and the Quality/New Goods Biases

By the early 2000s, the Bureau of Labor Statistics was using alternative mathematical methods for calculating the Consumer Price Index, more complicated than just adding up the cost of a fixed basket of goods, to allow for some substitution between goods. It was also updating the basket of goods behind the CPI more frequently, so that new and improved goods could be included more rapidly. For certain products, the BLS was carrying out studies to try to measure the quality improvement. For example, with computers, an economic study can try to adjust for changes in speed, memory, screen size, and other characteristics of the product, and then calculate the change in price after these product changes are taken into

account. But these adjustments are inevitably imperfect, and exactly how to make these adjustments is often a source of controversy among professional economists.

By the early 2000s, the substitution bias and quality/new goods bias had been somewhat reduced, so that since then the rise in the CPI probably overstates the true rise in inflation by only about 0.5% per year. Over one or a few years, this is not much; over a period of a decade or two, even half of a percent per year compounds to a more significant amount. In addition, the CPI tracks prices from physical locations, and not at online sites like Amazon, where prices can be lower.

When measuring inflation (and other economic statistics, too), a tradeoff arises between simplicity and interpretation. If the inflation rate is calculated with a basket of goods that is fixed and unchanging, then the calculation of an inflation rate is straightforward, but the problems of substitution bias and quality/new goods bias will arise. However, when the basket of goods is allowed to shift and evolve to reflect substitution toward lower relative prices, quality improvements, and new goods, the technical details of calculating the inflation rate grow more complex.

Key Concepts and Summary

Measuring price levels with a fixed basket of goods will always have two problems: the substitution bias, by which a fixed basket of goods does not allow for buying more of what is relatively less expensive and less of what is relatively more expensive; and the quality/new goods bias, by which a fixed basket cannot take into account improvements in quality and the advent of new goods. These problems can be reduced in degree—for example, by allowing the basket of goods to evolve over time—but they cannot be totally eliminated. The most commonly cited measure of inflation is the Consumer Price Index (CPI), which is based on a basket of goods representing what the typical consumer buys. The Core Inflation Index further breaks down the CPI by excluding volatile economic variables. Several price indices are not based on baskets of consumer goods. The GDP deflator is based on all the components of GDP. The Producer Price Index is based on prices of supplies and inputs bought by producers of goods and

services. An Employment Cost Index measures wage inflation in the labor market. An International Price Index is based on the prices of merchandise that is exported or imported.

Self-Check Questions

Exercise:

Problem:

[How Changes in the Cost of Living are Measured](#) introduced a number of different price indices. Which price index would be best to use to adjust your paycheck for inflation?

Solution:

Since the CPI measures the prices of the goods and services purchased by the typical urban consumer, it measures the prices of things that people buy with their paycheck. For that reason, the CPI would be the best price index to use for this purpose.

Exercise:

Problem:

The Consumer Price Index is subject to the substitution bias and the quality/new goods bias. Are the Producer Price Index and the GDP Deflator also subject to these biases? Why or why not?

Solution:

The PPI is subject to those biases for essentially the same reasons as the CPI is. The GDP deflator picks up prices of what is actually purchased that year, so there are no biases. That is the advantage of using the GDP deflator over the CPI.

Review Questions

Exercise:**Problem:**

Why does “substitution bias” arise if the inflation rate is calculated based on a fixed basket of goods?

Exercise:**Problem:**

Why does the “quality/new goods bias” arise if the inflation rate is calculated based on a fixed basket of goods?

Critical Thinking Question**Exercise:****Problem:**

Given the federal budget deficit in recent years, some economists have argued that by adjusting Social Security payments for inflation using the CPI, Social Security is overpaying recipients. What is the argument being made, and do you agree or disagree with it?

Exercise:**Problem:**

Why is the GDP deflator not an accurate measure of inflation as it impacts a household?

Exercise:**Problem:**

Imagine that the government statisticians who calculate the inflation rate have been updating the basic basket of goods once every 10 years, but now they decide to update it every five years. How will this change affect the amount of substitution bias and quality/new goods bias?

Exercise:

Problem:

Describe a situation, either a government policy situation, an economic problem, or a private sector situation, where using the CPI to convert from nominal to real would be more appropriate than using the GDP deflator.

Exercise:**Problem:**

Describe a situation, either a government policy situation, an economic problem, or a private sector situation, where using the GDP deflator to convert from nominal to real would be more appropriate than using the CPI.

References

Bernhard, Kent. "Pump Prices Jump Across U.S. after Katrina." *NBC News*, September 1, 2005. http://www.nbcnews.com/id/9146363/ns/business-local_business/t/pump-prices-jump-across-us-after-katrina/#.U00kRfk7um4.

Wynne, Mark A. "Core Inflation, A Review of Some Conceptual Issues." *Federal Reserve Bank of St. Louis*. p. 209. Accessed April 14, 2014. <http://research.stlouisfed.org/publications/review/08/05/part2/Wynne.pdf>

Glossary**Consumer Price Index (CPI)**

a measure of inflation calculated by U.S. government statisticians based on the price level from a fixed basket of goods and services that represents the purchases of the average consumer

core inflation index

a measure of inflation typically calculated by taking the CPI and excluding volatile economic variables such as food and energy prices

to better measure the underlying and persistent trend in long-term prices

Employment Cost Index

a measure of inflation based on wages paid in the labor market

GDP deflator

a measure of inflation based on the prices of all the components of GDP

International Price Index

a measure of inflation based on the prices of merchandise that is exported or imported

Producer Price Index (PPI)

a measure of inflation based on prices paid for supplies and inputs by producers of goods and services

quality/new goods bias

inflation calculated using a fixed basket of goods over time tends to overstate the true rise in cost of living, because it does not take into account improvements in the quality of existing goods or the invention of new goods

substitution bias

an inflation rate calculated using a fixed basket of goods over time tends to overstate the true rise in the cost of living, because it does not take into account that the person can substitute away from goods whose prices rise by a lot

How the U.S. and Other Countries Experience Inflation

By the end of this section, you will be able to:

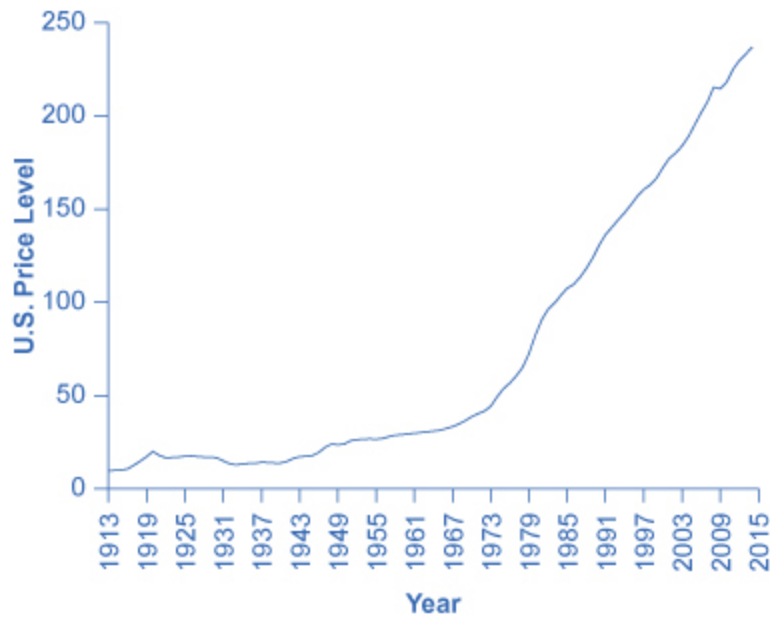
- Identify patterns of inflation for the United States using data from the Consumer Price Index
- Identify patterns of inflation on an international level

In the last three decades, inflation has been relatively low in the U.S. economy, with the Consumer Price Index typically rising 2% to 4% per year. Looking back over the twentieth century, there have been several periods where inflation caused the price level to rise at double-digit rates, but nothing has come close to hyperinflation.

Historical Inflation in the U.S. Economy

[\[link\]](#) (a) shows the level of prices in the Consumer Price Index stretching back to 1916. In this case, the base years (when the CPI is defined as 100) are set for the average level of prices that existed from 1982 to 1984. [\[link\]](#) (b) shows the annual percentage changes in the CPI over time, which is the inflation rate.

U.S. Price Level and Inflation Rates since 1913



(a) U.S. price level 1913-2014



(b) U.S. inflation rate 1913-2014

Graph a shows the trends in the U.S. price level from the year 1916 to 2014. In 1916, the graph starts out close to \$10, rises to around \$20 in 1920, stays around \$16 or \$17 until 1931, when it jumps to around \$15. It gradually increases, with periodic dips, until 2014, when it is around \$236. Graph b shows the trends in U.S. inflation rates from the year 1916 to 2014. In 1916, the graph starts out at 7.7%, jumps to close to

18% in 1917, drops drastically to close to –11% in 1921, goes up and down periodically, until settling to around 1.5% in 2014.

The first two waves of inflation are easy to characterize in historical terms: they are right after World War I and World War II. However, there are also two periods of severe negative inflation—called **deflation**—in the early decades of the twentieth century: one following the deep recession of 1920–21 and the other during the Great Depression of the 1930s. (Since inflation is a time when the buying power of money in terms of goods and services is reduced, deflation will be a time when the buying power of money in terms of goods and services increases.) For the period from 1900 to about 1960, the major inflations and deflations nearly balanced each other out, so the average annual rate of inflation over these years was only about 1% per year. A third wave of more severe inflation arrived in the 1970s and departed in the early 1980s.

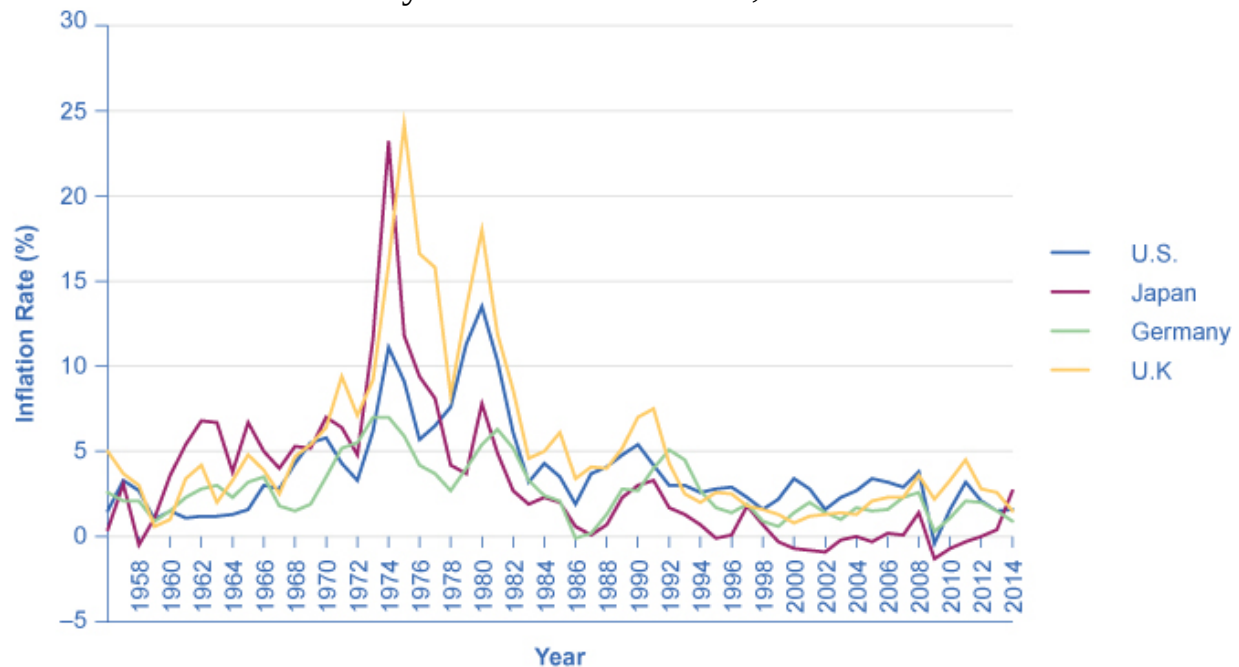
Note: Visit this [website](#) to use an inflation calculator and discover how prices have changed in the last 100 years.

Times of recession or depression often seem to be times when the inflation rate is lower, as in the recession of 1920–1921, the Great Depression, the recession of 1980–1982, and the Great Recession in 2008–2009. There were a few months in 2009 that were deflationary, but not at an annual rate. Recessions are typically accompanied by higher levels of unemployment, and the total demand for goods falls, pulling the price level down. Conversely, the rate of inflation often, but not always, seems to start moving up when the economy is growing very strongly, like right after wartime or during the 1960s. The frameworks for macroeconomic analysis, developed in other chapters, will explain why recession often accompanies higher unemployment and lower inflation, while rapid economic growth often brings lower unemployment but higher inflation.

Inflation around the World

Around the rest of the world, the pattern of inflation has been very mixed, as can be seen in [\[link\]](#) which shows inflation rates over the last several decades. Many industrialized countries, not just the United States, had relatively high inflation rates in the 1970s. For example, in 1975, Japan's inflation rate was over 8% and the inflation rate for the United Kingdom was almost 25%. In the 1980s, inflation rates came down in the United States and in Europe and have largely stayed down.

Countries with Relatively Low Inflation Rates, 1960–2014

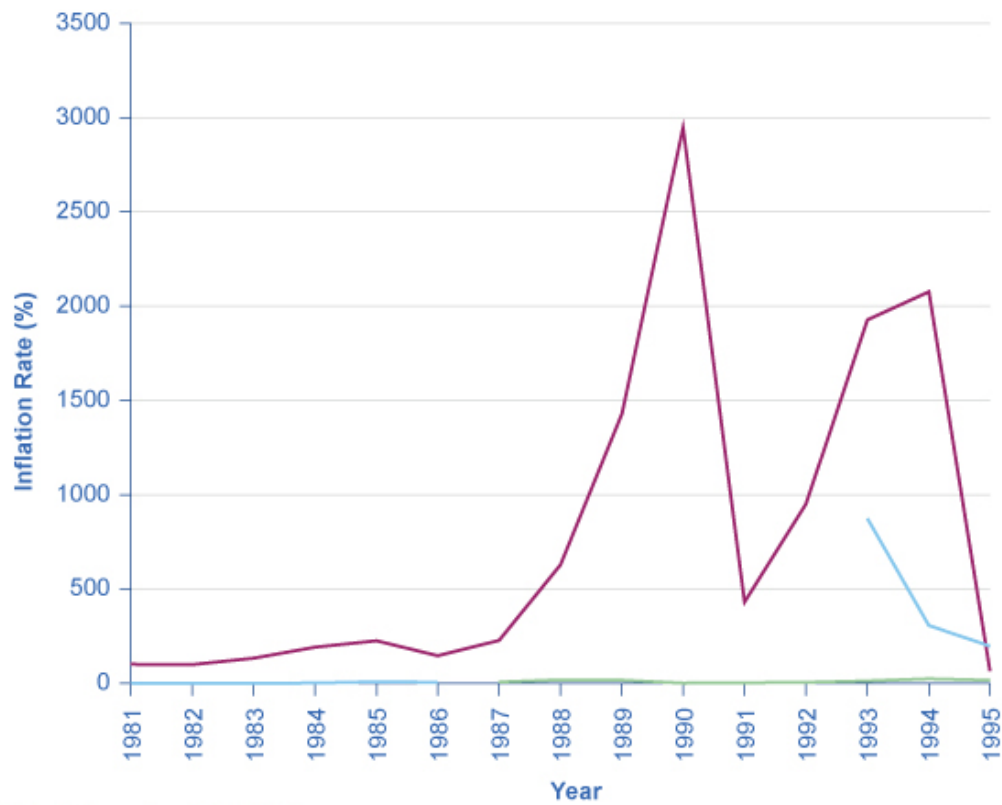


This chart shows the annual percentage change in consumer prices compared with the previous year's consumer prices in the United States, the United Kingdom, Japan, and Germany.

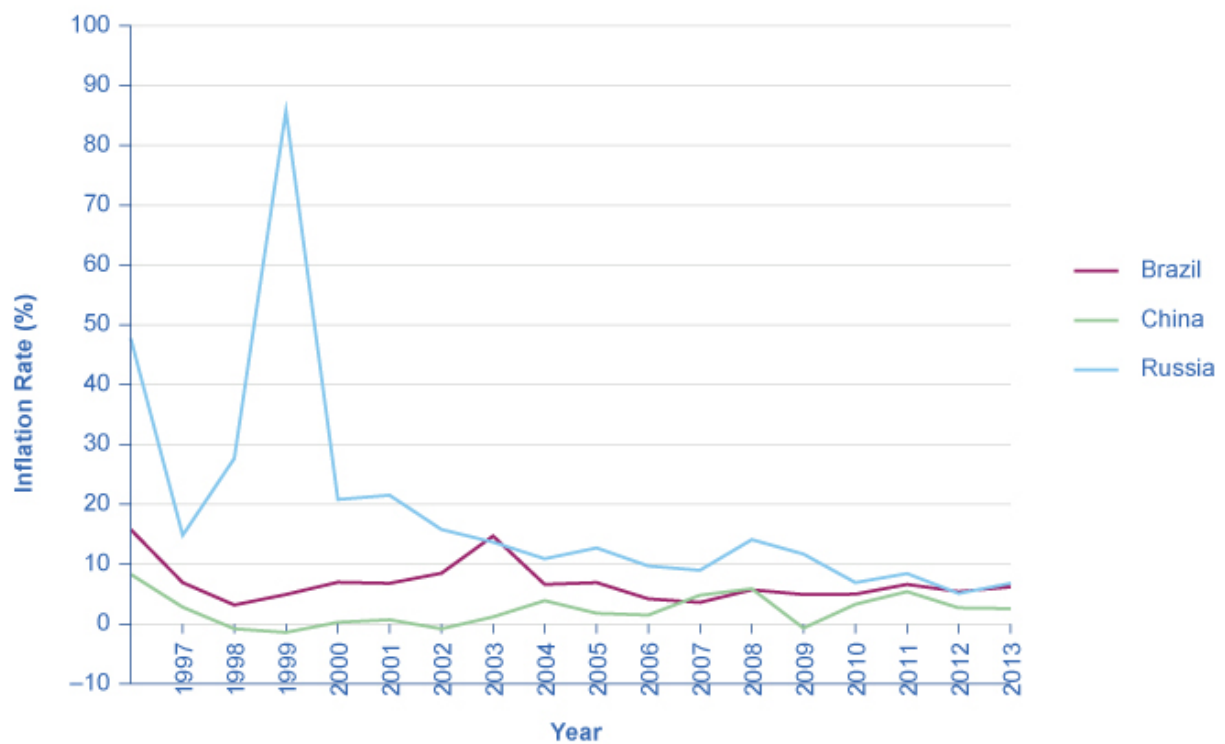
Countries with controlled economies in the 1970s, like the Soviet Union and China, historically had very low rates of measured inflation—because prices were forbidden to rise by law, except for the cases where the government deemed a price increase to be due to quality improvements. However, these countries also had perpetual shortages of goods, since

forbidding prices to rise acts like a price ceiling and creates a situation where quantity demanded often exceeds quantity supplied. As Russia and China made a transition toward more market-oriented economies, they also experienced outbursts of inflation, although the statistics for these economies should be regarded as somewhat shakier. Inflation in China averaged about 10% per year for much of the 1980s and early 1990s, although it has dropped off since then. Russia experienced **hyperinflation**—an outburst of high inflation—of 2,500% per year in the early 1990s, although by 2006 Russia's consumer price inflation had dipped below 10% per year, as shown in [\[link\]](#). The closest the United States has ever gotten to hyperinflation was during the Civil War, 1860–1865, in the Confederate states.

Countries with Relatively High Inflation Rates, 1980–2013



(a) Inflation rates 1980-1995



(b) Inflation rates 1996-2013

These charts show the percentage change in consumer prices compared with the previous year's consumer prices in Brazil, China, and Russia. (a) Of these, Brazil and Russia experienced hyperinflation at some point between the mid-1980s and mid-1990s. (b) Though not as high, China and Nigeria also had high inflation rates in the mid-1990s. Even though their inflation rates have come down over the last two decades, several of these countries continue to see significant inflation rates. (Sources:

<http://research.stlouisfed.org/fred2/series/FPCPITOTLZGBRA>;
<http://research.stlouisfed.org/fred2/series/CHNCPIALLMINMEI>;
<http://research.stlouisfed.org/fred2/series/FPCPITOTLZGRUS>)

Many countries in Latin America experienced raging hyperinflation during the 1980s and early 1990s, with inflation rates often well above 100% per year. In 1990, for example, both Brazil and Argentina saw inflation climb above 2000%. Certain countries in Africa experienced extremely high rates of inflation, sometimes bordering on hyperinflation, in the 1990s. Nigeria, the most populous country in Africa, had an inflation rate of 75% in 1995.

In the early 2000s, the problem of inflation appears to have diminished for most countries, at least in comparison to the worst times of recent decades. As we noted in this earlier Bring it Home feature, in recent years, the world's worst example of hyperinflation was in Zimbabwe, where at one point the government was issuing bills with a face value of \$100 trillion (in Zimbabwean dollars)—that is, the bills had \$100,000,000,000,000 written on the front, but were almost worthless. In many countries, the memory of double-digit, triple-digit, and even quadruple-digit inflation is not very far in the past.

Key Concepts and Summary

In the U.S. economy, the annual inflation rate in the last two decades has typically been around 2% to 4%. The periods of highest inflation in the United States in the twentieth century occurred during the years after World

Wars I and II, and in the 1970s. The period of lowest inflation—actually, with deflation—was the Great Depression of the 1930s.

Self-Check Question

Exercise:

Problem:

Go to this [website](#) for the Purchasing Power Calculator at MeasuringWorth.com. How much money would it take today to purchase what one dollar would have bought in the year of your birth?

Solution:

The calculator requires you to input three numbers:

- The first year, in this case the year of your birth
- The amount of money you would want to translate in terms of its purchasing power
- The last year—now or the most recent year the calculator will accept

My birth year is 1955. The amount is \$1. The year 2012 is currently the latest year the calculator will accept. The simple purchasing power calculator shows that \$1 of purchases in 1955 would cost \$8.57 in 2012. The website also explains how the true answer is more complicated than that shown by the simple purchasing power calculator.

Review Questions

Exercise:

Problem:

What has been a typical range of inflation in the U.S. economy in the last decade or so?

Exercise:**Problem:**

Over the last century, during what periods was the U.S. inflation rate highest and lowest?

Exercise:

Problem: What is deflation?

Critical Thinking Question**Exercise:****Problem:**

Why do you think the U.S. experience with inflation over the last 50 years has been so much milder than in many other countries?

Problems**Exercise:****Problem:**

Within 1 or 2 percentage points, what has the U.S. inflation rate been during the last 20 years? Draw a graph to show the data.

Glossary

deflation

negative inflation; most prices in the economy are falling

hyperinflation

an outburst of high inflation that is often seen (although not exclusively) when economies shift from a controlled economy to a

market-oriented economy

The Confusion Over Inflation

By the end of this section, you will be able to:

- Explain how inflation can cause redistributions of purchasing power
- Identify ways inflation can blur the perception of supply and demand
- Explain the economic benefits and challenges of inflation

Economists usually oppose high inflation, but they oppose it in a milder way than many non-economists. Robert Shiller, one of 2013's Nobel Prize winners in economics, carried out several surveys during the 1990s about attitudes toward inflation. One of his questions asked, "Do you agree that preventing high inflation is an important national priority, as important as preventing drug abuse or preventing deterioration in the quality of our schools?" Answers were on a scale of 1–5, where 1 meant "Fully agree" and 5 meant "Completely disagree." For the U.S. population as a whole, 52% answered "Fully agree" that preventing high inflation was a highly important national priority and just 4% said "Completely disagree." However, among professional economists, only 18% answered "Fully agree," while the same percentage of 18% answered "Completely disagree."

The Land of Funny Money

What are the economic problems caused by inflation, and why do economists often regard them with less concern than the general public? Consider a very short story: "The Land of Funny Money."

One morning, everyone in the Land of Funny Money awakened to find that everything denominated in money had increased by 20%. The change was completely unexpected. Every price in every store was 20% higher. Paychecks were 20% higher. Interest rates were 20 % higher. The amount of money, everywhere from wallets to savings accounts, was 20% larger. This overnight inflation of prices made newspaper headlines everywhere in the Land of Funny Money. But the headlines quickly disappeared, as people realized that in terms of what they could actually buy with their incomes, this inflation had no economic impact. Everyone's pay could still buy exactly the same set of goods as it did before. Everyone's savings were still

sufficient to buy exactly the same car, vacation, or retirement that they could have bought before. Equal levels of inflation in all wages and prices ended up not mattering much at all.

When the people in Robert Shiller's surveys explained their concern about inflation, one typical reason was that they feared that as prices rose, they would not be able to afford to buy as much. In other words, people were worried because they did not live in a place like the Land of Funny Money, where all prices and wages rose simultaneously. Instead, people live here on Planet Earth, where prices might rise while wages do not rise at all, or where wages rise more slowly than prices.

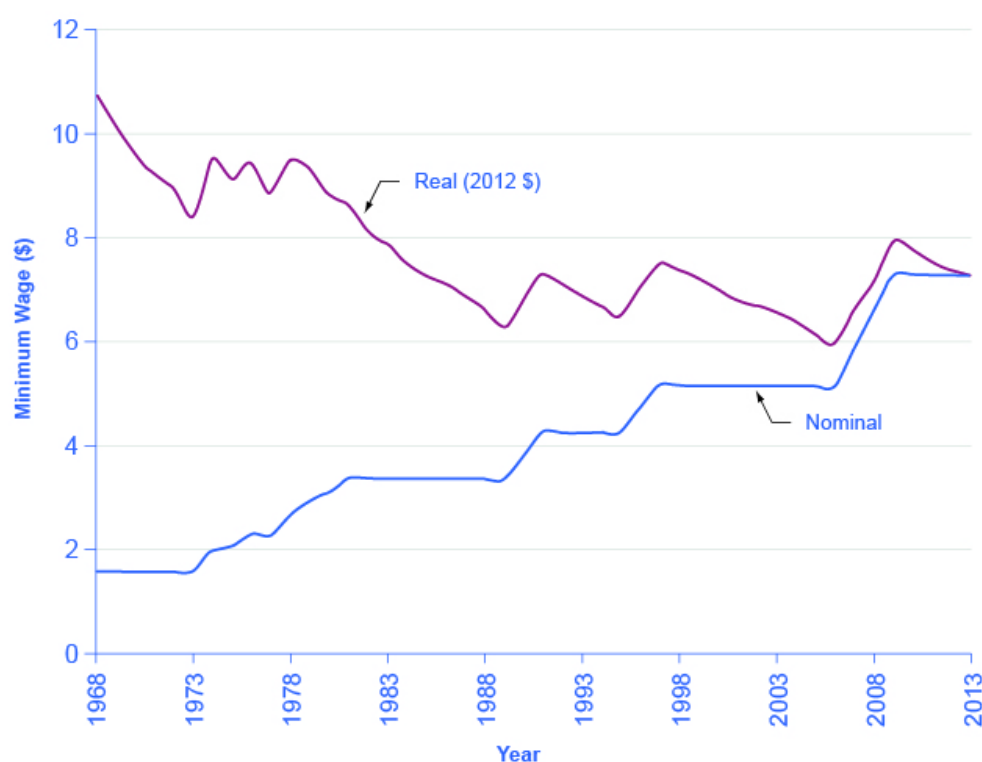
Economists note that over most periods, the inflation level in prices is roughly similar to the inflation level in wages, and so they reason that, on average, over time, people's economic status is not greatly changed by inflation. If all prices, wages, and interest rates adjusted automatically and immediately with inflation, as in the Land of Funny Money, then no one's purchasing power, profits, or real loan payments would change. However, if other economic variables do not move exactly in sync with inflation, or if they adjust for inflation only after a time lag, then inflation can cause three types of problems: unintended redistributions of purchasing power, blurred price signals, and difficulties in long-term planning.

Unintended Redistributions of Purchasing Power

Inflation can cause redistributions of purchasing power that hurt some and help others. People who are hurt by inflation include those who are holding a lot of cash, whether it is in a safe deposit box or in a cardboard box under the bed. When inflation happens, the buying power of cash is diminished. But cash is only an example of a more general problem: anyone who has financial assets invested in a way that the nominal return does not keep up with inflation will tend to suffer from inflation. For example, if a person has money in a bank account that pays 4% interest, but inflation rises to 5%, then the real rate of return for the money invested in that bank account is negative 1%.

Inflation can cause unintended redistributions for wage earners, too. Wages do typically creep up with inflation over time eventually. The last row of [\[link\]](#) at the start of this chapter showed that average hourly wage in the U.S. economy increased from \$3.23 in 1970 to \$19.55 in 2014, which is an increase by a factor of almost six. Over that time period, the Consumer Price Index increased by an almost identical amount. However, increases in wages may lag behind inflation for a year or two, since wage adjustments are often somewhat sticky and occur only once or twice a year. Moreover, the extent to which wages keep up with inflation creates insecurity for workers and may involve painful, prolonged conflicts between employers and employees. If the minimum wage is adjusted for inflation only infrequently, minimum wage workers are losing purchasing power from their nominal wages, as shown in [\[link\]](#).

U.S. Minimum Wage and Inflation



After adjusting for inflation, the federal minimum wage dropped more than 30 percent from 1967 to 2010, even though the nominal figure climbed from \$1.40 to \$7.25 per hour. Increases in the minimum wage in between 2008 and 2010 kept the decline from being worse—as it would have been if the wage had remained the same as it did from

1997 through 2007. (Sources:
<http://www.dol.gov/whd/minwage/chart.htm>; <http://data.bls.gov/cgi-bin/surveymost?cu>)

The unintended redistributions of buying power caused by inflation may have a broader effect on society. America's widespread acceptance of market forces rests on a perception that people's actions have a reasonable connection to market outcomes. When inflation causes a retiree who built up a pension or invested at a fixed interest rate to suffer, however, while someone who borrowed at a fixed interest rate benefits from inflation, it is hard to believe that this outcome was deserved in any way. Similarly, when homeowners benefit from inflation because the price of their homes rises, while renters suffer because they are paying higher rent, it is hard to see any useful incentive effects. One of the reasons that inflation is so disliked by the general public is a sense that it makes economic rewards and penalties more arbitrary—and therefore likely to be perceived as unfair – even dangerous, as the next Clear It Up feature shows.

Note:

Is there a connection between German hyperinflation and Hitler's rise to power?

Germany suffered an intense hyperinflation of its currency, the Mark, in the years after World War I, when the Weimar Republic in Germany resorted to printing money to pay its bills and the onset of the Great Depression created the social turmoil that Adolf Hitler could take advantage of in his rise to power. Shiller described the connection this way in a National Bureau of Economic Research 1996 Working Paper:

"A fact that is probably little known to young people today, even in Germany, is that the final collapse of the Mark in 1923, the time when the Mark's inflation reached astronomical levels (inflation of 35,974.9% in November 1923 alone, for an annual rate that month of $4.69 \times 10^{28}\%$), came in the same month as did Hitler's Beer Hall Putsch, his Nazi Party's armed attempt to overthrow the German government. This failed putsch

resulted in Hitler's imprisonment, at which time he wrote his book *Mein Kampf*, setting forth an inspirational plan for Germany's future, suggesting plans for world domination. . . " " . . . Most people in Germany today probably do not clearly remember these events; this lack of attention to it may be because its memory is blurred by the more dramatic events that succeeded it (the Nazi seizure of power and World War II). However, to someone living through these historical events in sequence . . . [the putsch] may have been remembered as vivid evidence of the potential effects of inflation."

Problems of Long-Term Planning

Inflation can make long-term planning difficult. In discussing unintended redistributions, we considered the case of someone trying to plan for retirement with a pension that is fixed in nominal terms and a high rate of inflation. Similar problems arise for all people trying to save for retirement, because they must consider what their money will really buy several decades in the future when the rate of future inflation cannot be known with certainty.

Inflation, especially at moderate or high levels, will pose substantial planning problems for businesses, too. A firm can make money from inflation—for example, by paying bills and wages as late as possible so that it can pay in inflated dollars, while collecting revenues as soon as possible. A firm can also suffer losses from inflation, as in the case of a retail business that gets stuck holding too much cash, only to see the value of that cash eroded by inflation. But when a business spends its time focusing on how to profit by inflation, or at least how to avoid suffering from it, an inevitable tradeoff strikes: less time is spent on improving products and services or on figuring out how to make existing products and services more cheaply. An economy with high inflation rewards businesses that have found clever ways of profiting from inflation, which are not necessarily the businesses that excel at productivity, innovation, or quality of service.

Any Benefits of Inflation?

Although the economic effects of inflation are primarily negative, two countervailing points are worth noting. First, the impact of inflation will differ considerably according to whether it is creeping up slowly at 0% to 2% per year, galloping along at 10% to 20% per year, or racing to the point of hyperinflation at, say, 40% per month. Hyperinflation can rip an economy and a society apart. An annual inflation rate of 2%, 3%, or 4%, however, is a long way from a national crisis. Low inflation is also better than deflation which occurs with severe recessions.

Second, an argument is sometimes made that moderate inflation may help the economy by making wages in labor markets more flexible. The discussion in [Unemployment](#) pointed out that wages tend to be sticky in their downward movements and that unemployment can result. A little inflation could nibble away at real wages, and thus help real wages to decline if necessary. In this way, even if a moderate or high rate of inflation may act as sand in the gears of the economy, perhaps a low rate of inflation serves as oil for the gears of the labor market. This argument is controversial. A full analysis would have to take all the effects of inflation into account. It does, however, offer another reason to believe that, all things considered, very low rates of inflation may not be especially harmful.

Key Concepts and Summary

Unexpected inflation will tend to hurt those whose money received, in terms of wages and interest payments, does not rise with inflation. In contrast, inflation can help those who owe money that can be paid in less valuable, inflated dollars. Low rates of inflation have relatively little economic impact over the short term. Over the medium and the long term, even low rates of inflation can complicate future planning. High rates of inflation can muddle price signals in the short term and prevent market forces from operating efficiently, and can vastly complicate long-term savings and investment decisions.

Self-Check Question

Exercise:

Problem:

If inflation rises unexpectedly by 5%, would a state government that had recently borrowed money to pay for a new highway benefit or lose?

Solution:

The state government would benefit because it would repay the loan in less valuable dollars than it borrowed. Plus, tax revenues for the state government would increase because of the inflation.

Review Question**Exercise:****Problem:**

Identify several parties likely to be helped and hurt by inflation.

Critical Thinking Questions**Exercise:****Problem:**

If, over time, wages and salaries on average rise at least as fast as inflation, why do people worry about how inflation affects incomes?

Exercise:

Problem: Who in an economy is the big winner from inflation?

References

Shiller, Robert. “Why Do People Dislike Inflation?” *NBER Working Paper Series, National Bureau of Economic Research*, p. 52. 1996.

Indexing and Its Limitations

By the end of this section, you will be able to:

- Explain the relationship between indexing and inflation
- Identify three ways the government can control inflation through macroeconomic policy

When a price, wage, or interest rate is adjusted automatically with inflation, it is said to be **indexed**. An indexed payment increases according to the index number that measures inflation. A wide array of indexing arrangements is observed in private markets and government programs. Since the negative effects of inflation depend in large part on having inflation unexpectedly affect one part of the economy but not another—say, increasing the prices that people pay but not the wages that workers receive—indexing will take some of the sting out of inflation.

Indexing in Private Markets

In the 1970s and 1980s, labor unions commonly negotiated wage contracts that had **cost-of-living adjustments (COLAs)** which guaranteed that their wages would keep up with inflation. These contracts were sometimes written as, for example, COLA plus 3%. Thus, if inflation was 5%, the wage increase would automatically be 8%, but if inflation rose to 9%, the wage increase would automatically be 12%. COLAs are a form of indexing applied to wages.

Loans often have built-in inflation adjustments, too, so that if the inflation rate rises by two percentage points, then the interest rate charged on the loan rises by two percentage points as well. An **adjustable-rate mortgage (ARM)** is a kind of loan used to purchase a home in which the interest rate varies with the rate of inflation. Often, a borrower will be able receive a lower interest rate if borrowing with an ARM, compared to a fixed-rate loan. The reason is that with an ARM, the lender is protected against the risk that higher inflation will reduce the real loan payments, and so the risk premium part of the interest rate can be correspondingly lower.

A number of ongoing or long-term business contracts also have provisions that prices will be adjusted automatically according to inflation. Sellers like such contracts because they are not locked into a low nominal selling price if inflation turns out higher than expected; buyers like such contracts because they are not locked into a high buying price if inflation turns out to be lower than expected. A contract with automatic adjustments for inflation in effect agrees on a real price to be paid, rather than a nominal price.

Indexing in Government Programs

The Social Security program offers two examples of indexing. Since the passage of the Social Security Indexing Act of 1972, the level of Social Security benefits increases each year along with the Consumer Price Index. Also, Social Security is funded by payroll taxes, which are imposed on the income earned up to a certain amount—\$117,000 in 2014. This level of income is adjusted upward each year according to the rate of inflation, so that the indexed rise in the benefit level is accompanied by an indexed increase in the Social Security tax base.

A Preview of Policy Discussions of Inflation

This chapter has focused on how inflation is measured, historical experience with inflation, how to adjust nominal variables into real ones, how inflation affects the economy, and how indexing works. The causes of inflation have barely been hinted at, and government policies to deal with inflation have not been addressed at all. These issues will be taken up in depth in other chapters. However, it is useful to offer a preview here.

The cause of inflation can be summed up in one sentence: Too many dollars chasing too few goods. The great surges of inflation early in the twentieth century came after wars, which are a time when government spending is very high, but consumers have little to buy, because production is going to the war effort. Governments also commonly impose price controls during wartime. After the war, the price controls end and pent-up buying power surges forth, driving up inflation. On the other hand, if too few dollars are chasing too many goods, then inflation will decline or even turn into deflation. Therefore, slowdowns in economic activity, as in major

recessions and the Great Depression, are typically associated with a reduction in inflation or even outright deflation.

The policy implications are clear. If inflation is to be avoided, the amount of purchasing power in the economy must grow at roughly the same rate as the production of goods. Macroeconomic policies that the government can use to affect the amount of purchasing power—through taxes, spending, and regulation of interest rates and credit—can thus cause inflation to rise or reduce inflation to lower levels.

Note:

A \$550 Million Loaf of Bread?

As we will learn in [Money and Banking](#), the existence of money provides enormous benefits to an economy. In a real sense, money is the lubrication that enhances the workings of markets. Money makes transactions easier. It allows people to find employment producing one product, then use the money earned to purchase the other products they need to live on.

However, too much money in circulation can lead to inflation. Extreme cases of governments recklessly printing money lead to hyperinflation. Inflation reduces the value of money. Hyperinflation, because money loses value so quickly, ultimately results in people no longer using money. The economy reverts to barter, or it adopts another country's more stable currency, like U.S. dollars. In the meantime, the economy literally falls apart as people leave jobs and fend for themselves because it is not worth the time to work for money that will be worthless in a few days.

Only national governments have the power to cause hyperinflation.

Hyperinflation typically happens when government faces extraordinary demands for spending, which it cannot finance by taxes or borrowing. The only option is to print money—more and more of it. With more money in circulation chasing the same amount (or even less) goods and services, the only result is higher and higher prices until the economy and/or the government collapses. This is why economists are generally wary of letting inflation get out of control.

Key Concepts and Summary

A payment is said to be indexed if it is automatically adjusted for inflation. Examples of indexing in the private sector include wage contracts with cost-of-living adjustments (COLAs) and loan agreements like adjustable-rate mortgages (ARMs). Examples of indexing in the public sector include tax brackets and Social Security payments.

Self-Check Questions

Exercise:

Problem:

How should an increase in inflation affect the interest rate on an adjustable-rate mortgage?

Solution:

Higher inflation reduces real interest rates on fixed rate mortgages. Because ARMs can be adjusted, higher inflation leads to higher interest rates on ARMs.

Exercise:

Problem:

A fixed-rate mortgage has the same interest rate over the life of the loan, whether the mortgage is for 15 or 30 years. By contrast, an adjustable-rate mortgage changes with market interest rates over the life of the mortgage. If inflation falls unexpectedly by 3%, what would likely happen to a homeowner with an adjustable-rate mortgage?

Solution:

Because the mortgage has an adjustable rate, the rate should fall by 3%, the same as inflation, to keep the real interest rate the same.

Review Questions

Exercise:

Problem: What is indexing?

Exercise:

Problem:

Name several forms of indexing in the private and public sector.

Critical Thinking Questions

Exercise:

Problem:

If a government gains from unexpected inflation when it borrows, why would it choose to offer indexed bonds?

Exercise:

Problem: Do you think perfect indexing is possible? Why or why not?

Problems

Exercise:

Problem:

If inflation rises unexpectedly by 5%, indicate for each of the following whether the economic actor is helped, hurt, or unaffected:

- a. A union member with a COLA wage contract
- b. Someone with a large stash of cash in a safe deposit box
- c. A bank lending money at a fixed rate of interest

- d. A person who is not due to receive a pay raise for another 11 months

Exercise:

Problem:

Rosalie the Retiree knows that when she retires in 16 years, her company will give her a one-time payment of \$20,000. However, if the inflation rate is 6% per year, how much buying power will that \$20,000 have when measured in today's dollars? *Hint:* Start by calculating the rise in the price level over the 16 years.

References

Wines, Michael. "How Bad is Inflation in Zimbabwe?" *The New York Times*, May 2, 2006.

http://www.nytimes.com/2006/05/02/world/africa/02zimbabwe.html?pagewanted=all&_r=0.

Hanke, Steve H. "R.I.P. Zimbabwe Dollar." *CATO Institute*. Accessed December 31, 2013. <http://www.cato.org/zimbabwe>.

Massachusetts Institute of Technology. 2015. "Billion Prices Project." Accessed March 4, 2015. <http://bpp.mit.edu/usa/>.

Glossary

adjustable-rate mortgage (ARM)

a loan used to purchase a home in which the interest rate varies with market interest rates

cost-of-living adjustments (COLAs)

a contractual provision that wage increases will keep up with inflation

indexed

a price, wage, or interest rate is adjusted automatically for inflation

Introduction to Government Budgets and Fiscal Policy

class="introduction"

Shut Downs and Parks

Yellowstone National Park is one of the many national parks forced to close down during the government shut down in October 2013. (Credit: modification of work by “daveynin”/flickr r Creative Commons)



Note:**No Yellowstone Park?**

So you had trekked all the way to see Yellowstone National Park in the beautiful month of October 2013, only to find it... closed. Closed! Why? For two weeks in October 2013, the U.S. federal government shut down. Many federal services, like the national parks, closed and 800,000 federal employees were furloughed. Tourists were shocked and so was the rest of the world: Congress and the President could not agree on a budget. Inside the Capitol, Republicans and Democrats argued about spending priorities and whether to increase the national debt limit. Each year's budget, which is over \$3 trillion of spending, must be approved by Congress and signed by the President. Two thirds of the budget is entitlements and other mandatory spending which occur without congressional or presidential action once the programs are set up. Tied to the budget debate was the issue of increasing the debt ceiling—how high the national debt of the U.S. government can be. The House of Representatives refused to sign on to the bills to fund the government unless they included provisions to stop or change the Affordable Health Care Act (more colloquially known as Obamacare). As the days ticked by, the United States came very close to defaulting on its debt.

Why does the federal budget create such intense debates? What would happen if the United States actually defaulted on its debt? In this chapter, we will examine the federal budget, taxation, and fiscal policy. We will also look at the annual federal budget deficits and the national debt.

Note:**Introduction to Government Budgets and Fiscal Policy**

In this chapter, you will learn about:

- Government Spending
- Taxation
- Federal Deficits and the National Debt
- Using Fiscal Policy to Fight Recessions, Unemployment, and Inflation
- Automatic Stabilizers

- Practical Problems with Discretionary Fiscal Policy
- The Question of a Balanced Budget

All levels of government—federal, state, and local—have budgets that show how much revenue the government expects to receive in taxes and other income and how the government plans to spend it. Budgets, however, can shift dramatically within a few years, as policy decisions and unexpected events shake up earlier tax and spending plans.

In this chapter, we revisit fiscal policy, which was first covered in [Welcome to Economics!](#) Fiscal policy is one of two policy tools for fine tuning the economy (the other is monetary policy). While monetary policy is made by policymakers at the Federal Reserve, fiscal policy is made by Congress and the President.

The discussion of fiscal policy focuses on how federal government taxing and spending affects aggregate demand. All government spending and taxes affect the economy, but fiscal policy focuses strictly on the policies of the federal government. We begin with an overview of U.S. government spending and taxes. We then discuss fiscal policy from a short-run perspective; that is, how government uses tax and spending policies to address recession, unemployment, and inflation; how periods of recession and growth affect government budgets; and the merits of balanced budget proposals.

Government Spending

By the end of this section, you will be able to:

- Identify U.S. budget deficit and surplus trends over the past five decades
- Explain the differences between the U.S. federal budget, and state and local budgets

Government spending covers a range of services provided by the federal, state, and local governments. When the federal government spends more money than it receives in taxes in a given year, it runs a **budget deficit**. Conversely, when the government receives more money in taxes than it spends in a year, it runs a **budget surplus**. If government spending and taxes are equal, it is said to have a **balanced budget**. For example, in 2009, the U.S. government experienced its largest budget deficit ever, as the federal government spent \$1.4 trillion more than it collected in taxes. This deficit was about 10% of the size of the U.S. GDP in 2009, making it by far the largest budget deficit relative to GDP since the mammoth borrowing used to finance World War II.

This section presents an overview of government spending in the United States.

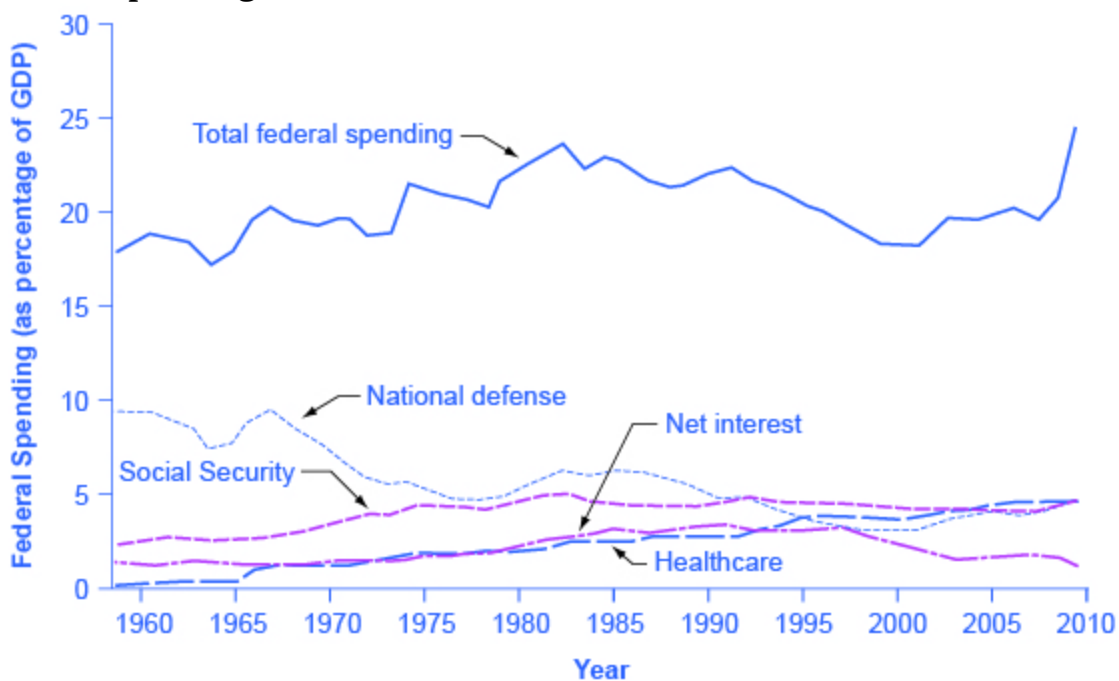
Total U.S. Government Spending

Federal spending in nominal dollars (that is, dollars not adjusted for inflation) has grown by a multiple of more than 38 over the last four decades, from \$93.4 billion in 1960 to \$3.9 trillion in 2014. Comparing spending over time in nominal dollars is misleading because it does not take into account inflation or growth in population and the real economy. A more useful method of comparison is to examine government spending as a percent of GDP over time.

The top line in [\[link\]](#) shows the level of federal spending since 1960, expressed as a share of GDP. Despite a widespread sense among many Americans that the federal government has been growing steadily larger, the

graph shows that federal spending has hovered in a range from 18% to 22% of GDP most of the time since 1960. The other lines in [\[link\]](#) show the major federal spending categories: national defense, Social Security, health programs, and interest payments. From the graph, we see that national defense spending as a share of GDP has generally declined since the 1960s, although there were some upward bumps in the 1980s buildup under President Ronald Reagan and in the aftermath of the terrorist attacks on September 11, 2001. In contrast, Social Security and healthcare have grown steadily as a percent of GDP. Healthcare expenditures include both payments for senior citizens (Medicare), and payments for low-income Americans (Medicaid). Medicaid is also partially funded by state governments. Interest payments are the final main category of government spending shown in the figure.

Federal Spending, 1960–2014



Since 1960, total federal spending has ranged from about 18% to 22% of GDP, although it climbed above that level in 2009, but quickly dropped back down to that level by 2013. The share spent on national defense has generally declined, while the share spent on Social Security and on healthcare expenses (mainly Medicare and Medicaid) has increased. (Source: *Economic Report of the*

President, Tables B-2 and B-22,
<http://www.gpo.gov/fdsys/pkg/ERP-2014/content-detail.html>)

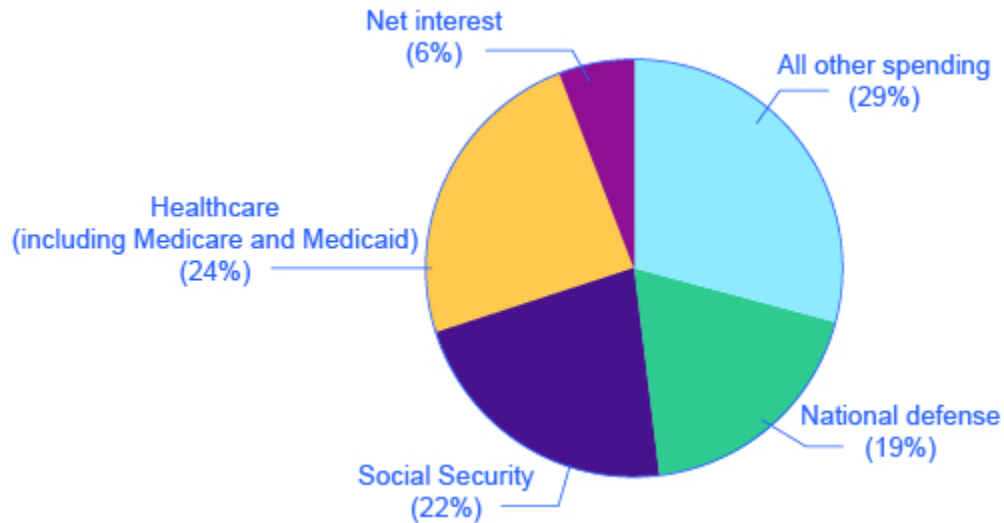
Each year, the government borrows funds from U.S. citizens and foreigners to cover its budget deficits. It does this by selling securities (Treasury bonds, notes, and bills)—in essence borrowing from the public and promising to repay with interest in the future. From 1961 to 1997, the U.S. government has run budget deficits, and thus borrowed funds, in almost every year. It had budget surpluses from 1998 to 2001, and then returned to deficits.

The interest payments on past federal government borrowing were typically 1–2% of GDP in the 1960s and 1970s but then climbed above 3% of GDP in the 1980s and stayed there until the late 1990s. The government was able to repay some of its past borrowing by running surpluses from 1998 to 2001 and, with help from low interest rates, the interest payments on past federal government borrowing had fallen back to 1.4% of GDP by 2012.

We investigate the patterns of government borrowing and debt in more detail later in this chapter, but first we need to clarify the difference between the deficit and the debt. *The deficit is not the debt.* The difference between the deficit and the debt lies in the time frame. The government deficit (or surplus) refers to what happens with the federal government budget each year. The government debt is accumulated over time; it is the sum of all past deficits and surpluses. If you borrow \$10,000 per year for each of the four years of college, you might say that your annual deficit was \$10,000, but your accumulated debt over the four years is \$40,000.

These four categories—national defense, Social Security, healthcare, and interest payments—account for roughly 73% of all federal spending, as [\[link\]](#) shows. The remaining 27% wedge of the pie chart covers all other categories of federal government spending: international affairs; science and technology; natural resources and the environment; transportation; housing; education; income support for the poor; community and regional development; law enforcement and the judicial system; and the administrative costs of running the government.

Slices of Federal Spending, 2014

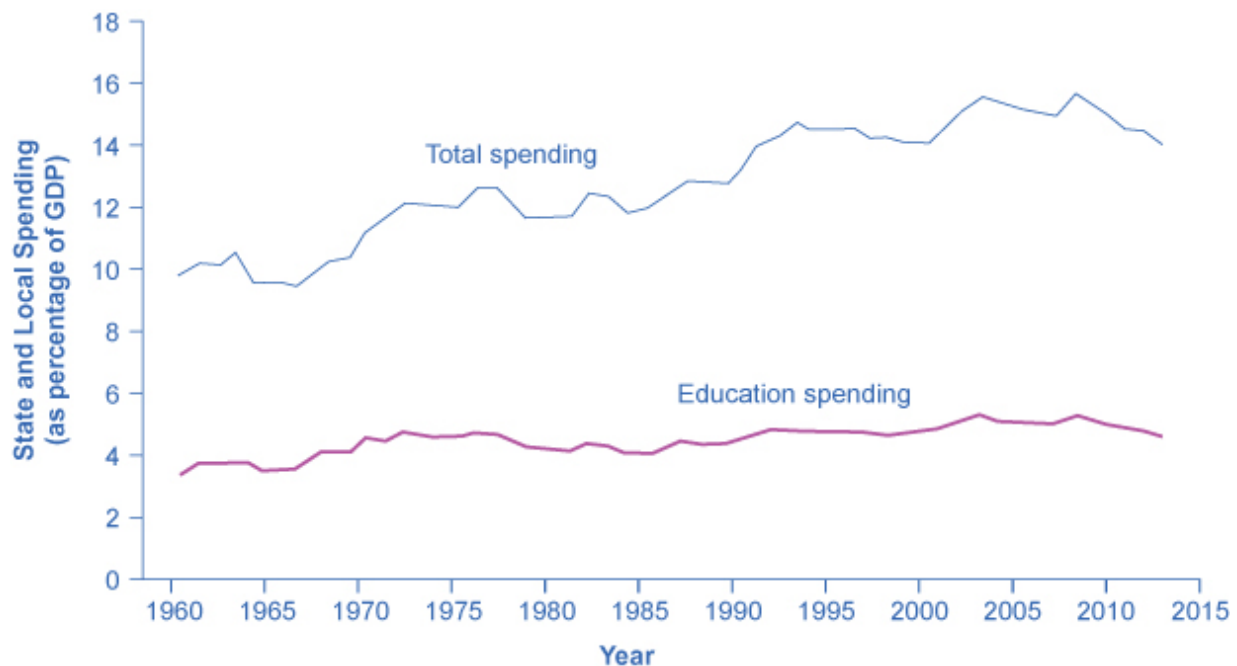


About 73% of government spending goes to four major areas: national defense, Social Security, healthcare, and interest payments on past borrowing. This leaves about 29% of federal spending for all other functions of the U.S. government. (Source: <https://www.whitehouse.gov/omb/budget/Historicals/>)

State and Local Government Spending

Although federal government spending often gets most of the media attention, state and local government spending is also substantial—at about \$3.1 trillion in 2014. [\[link\]](#) shows that state and local government spending has increased during the last four decades from around 8% to around 14% today. The single biggest item is education, which accounts for about one-third of the total. The rest covers programs like highways, libraries, hospitals and healthcare, parks, and police and fire protection. Unlike the federal government, all states (except Vermont) have balanced budget laws, which means any gaps between revenues and spending must be closed by higher taxes, lower spending, drawing down their previous savings, or some combination of all of these.

State and Local Spending, 1960–2013



Spending by state and local government increased from about 10% of GDP in the early 1960s to 14–16% by the mid-1970s. It has remained at roughly that level since. The single biggest spending item is education, including both K–12 spending and support for public colleges and universities, which has been about 4–5% of GDP in recent decades. Source: (Source: Bureau of Economic Analysis.)

U.S. presidential candidates often run for office pledging to improve the public schools or to get tough on crime. However, in the U.S. system of government, these tasks are primarily the responsibilities of state and local governments. Indeed, in fiscal year 2014 state and local governments spent about \$840 billion per year on education (including K–12 and college and university education), compared to only \$100 billion by the federal government, according to usgovernmentspending.com. In other words, about 90 cents of every dollar spent on education happens at the state and local level. A politician who really wants hands-on responsibility for reforming education or reducing crime might do better to run for mayor of a large city or for state governor rather than for president of the United States.

Key Concepts and Summary

Fiscal policy is the set of policies that relate to federal government spending, taxation, and borrowing. In recent decades, the level of federal government spending and taxes, expressed as a share of GDP, has not changed much, typically fluctuating between about 18% to 22% of GDP. However, the level of state spending and taxes, as a share of GDP, has risen from about 12–13% to about 20% of GDP over the last four decades. The four main areas of federal spending are national defense, Social Security, healthcare, and interest payments, which together account for about 70% of all federal spending. When a government spends more than it collects in taxes, it is said to have a budget deficit. When a government collects more in taxes than it spends, it is said to have a budget surplus. If government spending and taxes are equal, it is said to have a balanced budget. The sum of all past deficits and surpluses make up the government debt.

Self-Check Questions

Exercise:

Problem:

When governments run budget deficits, how do they make up the differences between tax revenue and spending?

Solution:

The government borrows funds by selling Treasury bonds, notes, and bills.

Exercise:

Problem:

When governments run budget surpluses, what is done with the extra funds?

Solution:

The funds can be used to pay down the national debt or else be refunded to the taxpayers.

Exercise:

Problem:

Is it possible for a nation to run budget deficits and still have its debt/GDP ratio fall? Explain your answer. Is it possible for a nation to run budget surpluses and still have its debt/GDP ratio rise? Explain your answer.

Solution:

Yes, a nation can run budget deficits and see its debt/GDP ratio fall. In fact, this is not uncommon. If the deficit is small in a given year, then the addition to debt in the numerator of the debt/GDP ratio will be relatively small, while the growth in GDP is larger, and so the debt/GDP ratio declines. This was the experience of the U.S. economy for the period from the end of World War II to about 1980. It is also theoretically possible, although not likely, for a nation to have a budget surplus and see its debt/GDP ratio rise. Imagine the case of a nation with a small surplus, but in a recession year when the economy shrinks. It is possible that the decline in the nation's debt, in the numerator of the debt/GDP ratio, would be proportionally less than the fall in the size of GDP, so the debt/GDP ratio would rise.

Review Questions

Exercise:

Problem:

Give some examples of changes in federal spending and taxes by the government that would be fiscal policy and some that would not.

Exercise:

Problem:

Have the spending and taxes of the U.S. federal government generally had an upward or a downward trend in the last few decades?

Exercise:**Problem:**

What are the main categories of U.S. federal government spending?

Exercise:**Problem:**

What is the difference between a budget deficit, a balanced budget, and a budget surplus?

Exercise:**Problem:**

Have spending and taxes by state and local governments in the United States had a generally upward or downward trend in the last few decades?

Critical Thinking Questions**Exercise:****Problem:**

Why is government spending typically measured as a percentage of GDP rather than in nominal dollars?

Exercise:**Problem:**

Why are expenditures such as crime prevention and education typically done at the state and local level rather than at the federal level?

Exercise:

Problem:

Why is spending by the U.S. government on scientific research at NASA fiscal policy while spending by the University of Illinois is not fiscal policy? Why is a cut in the payroll tax fiscal policy whereas a cut in a state income tax is not fiscal policy?

Problems

Exercise:

Problem:

A government starts off with a total debt of \$3.5 billion. In year one, the government runs a deficit of \$400 million. In year two, the government runs a deficit of \$1 billion. In year three, the government runs a surplus of \$200 million. What is the total debt of the government at the end of year three?

References

Kramer, Mattea, et. al. *A People's Guide to the Federal Budget*. National Priorities Project. Northampton: Interlink Books, 2012.

Kurtzleben, Danielle. "10 States With The Largest Budget Shortfalls." *U.S. News & World Report*. January 14, 2011.
<http://www.usnews.com/news/articles/2011/01/14/10-states-with-the-largest-budget-shortfalls>.

Miller, Rich, and William Selway. "U.S. Cities and States Start Spending Again." *BloombergBusinessweek*, January 10, 2013.
<http://www.businessweek.com/articles/2013-01-10/u-dot-s-dot-cities-and-states-start-spending-again>.

Weisman, Jonathan. "After Year of Working Around Federal Cuts, Agencies Face Fewer Options." *The New York Times*, October 26, 2013.
http://www.nytimes.com/2013/10/27/us/politics/after-year-of-working-around-federal-cuts-agencies-face-fewer-options.html?_r=0.

Chantrill, Christopher. USGovernmentSpending.com. "Government Spending Details: United States Federal State and Local Government Spending, Fiscal Year 2013."
http://www.usgovernmentspending.com/year_spending_2013USbn_15bs2n_20.

Glossary

balanced budget

when government spending and taxes are equal

budget deficit

when the federal government spends more money than it receives in taxes in a given year

budget surplus

when the government receives more money in taxes than it spends in a year

Taxation

By the end of this section, you will be able to:

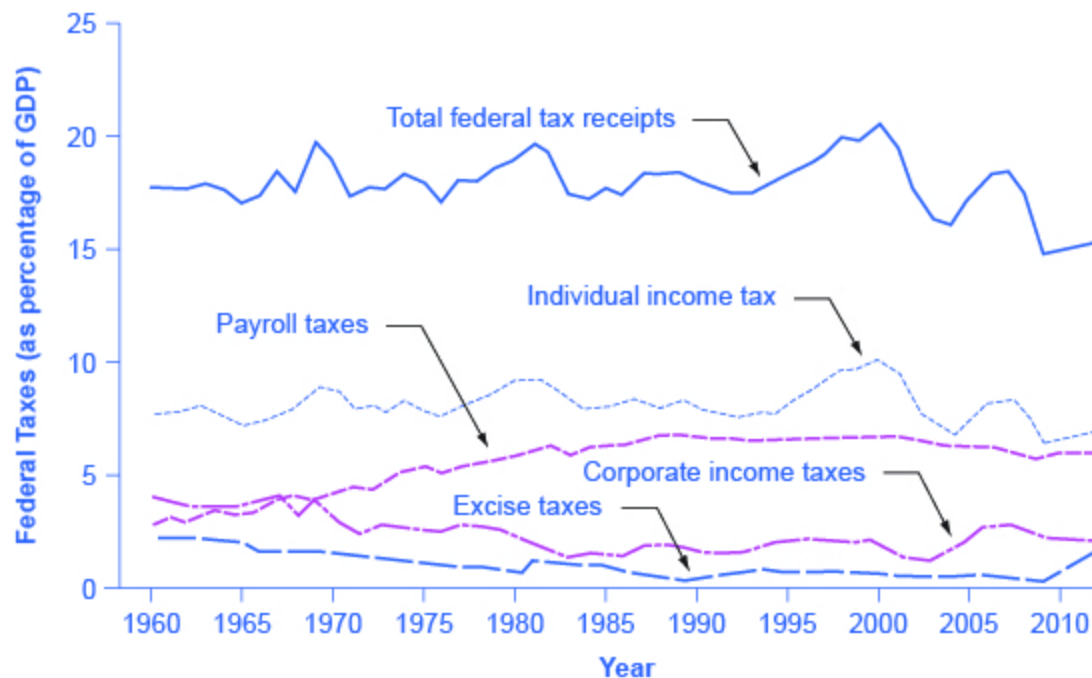
- Differentiate among a regressive tax, a proportional tax, and a progressive tax
- Identify the major sources of revenue for the U.S. federal budget

There are two main categories of taxes: those collected by the federal government and those collected by state and local governments. What percentage is collected and what that revenue is used for varies greatly. The following sections will briefly explain the taxation system in the United States.

Federal Taxes

Just as many Americans erroneously think that federal spending has grown considerably, many also believe that taxes have increased substantially. The top line of [\[link\]](#) shows total federal taxes as a share of GDP since 1960. Although the line rises and falls, it typically remains within the range of 17% to 20% of GDP, except for 2009, when taxes fell substantially below this level, due to recession.

Federal Taxes, 1960–2014



Federal tax revenues have been about 17–20% of GDP during most periods in recent decades. The primary sources of federal taxes are individual income taxes and the payroll taxes that finance Social Security and Medicare. Corporate income taxes and social insurance taxes provide smaller shares of revenue. (Source: *Economic Report of the President, 2015*. Table B-21, <https://www.whitehouse.gov/administration/eop/cea/economic-report-of-the-President/2015>)

[\[link\]](#) also shows the patterns of taxation for the main categories of taxes levied by the federal government: individual income taxes, corporate income taxes, and social insurance and retirement receipts. When most people think of taxes levied by the federal government, the first tax that comes to mind is the **individual income tax** that is due every year on April 15 (or the first business day after). The personal income tax is the largest single source of federal government revenue, but it still represents less than half of federal tax revenue.

The second largest source of federal revenue is the **payroll tax** (captured in social insurance and retirement receipts), which provides funds for Social Security and Medicare. Payroll taxes have increased steadily over time. Together, the personal income tax and the payroll tax accounted for about 80% of federal tax revenues in 2014. Although personal income tax revenues account for more total revenue than the payroll tax, nearly three-quarters of households pay more in payroll taxes than in income taxes.

The income tax is a **progressive tax**, which means that the tax rates increase as a household's income increases. Taxes also vary with marital status, family size, and other factors. The **marginal tax rates** (the tax that must be paid on all yearly income) for a single taxpayer range from 10% to 35%, depending on income, as the following Clear It Up feature explains.

Note:

How does the marginal rate work?

Suppose that a single taxpayer's income is \$35,000 per year. Also suppose that income from \$0 to \$9,075 is taxed at 10%, income from \$9,075 to \$36,900 is taxed at 15%, and, finally, income from \$36,900 and beyond is taxed at 25%. Since this person earns \$35,000, their marginal tax rate is 15%.

The key fact here is that the federal income tax is designed so that tax rates increase as income increases, up to a certain level. The payroll taxes that support Social Security and Medicare are designed in a different way. First, the payroll taxes for Social Security are imposed at a rate of 12.4% up to a certain wage limit, set at \$118,500 in 2015. Medicare, on the other hand, pays for elderly healthcare, and is fixed at 2.9%, with no upper ceiling.

In both cases, the employer and the employee split the payroll taxes. An employee only sees 6.2% deducted from his paycheck for Social Security, and 1.45% from Medicare. However, as economists are quick to point out, the employer's half of the taxes are probably passed along to the employees

in the form of lower wages, so in reality, the worker pays all of the payroll taxes.

The Medicare payroll tax is also called a **proportional tax**; that is, a flat percentage of all wages earned. The Social Security payroll tax is proportional up to the wage limit, but above that level it becomes a **regressive tax**, meaning that people with higher incomes pay a smaller share of their income in tax.

The third-largest source of federal tax revenue, as shown in [\[link\]](#) is the **corporate income tax**. The common name for corporate income is “profits.” Over time, corporate income tax receipts have declined as a share of GDP, from about 4% in the 1960s to an average of 1% to 2% of GDP in the first decade of the 2000s.

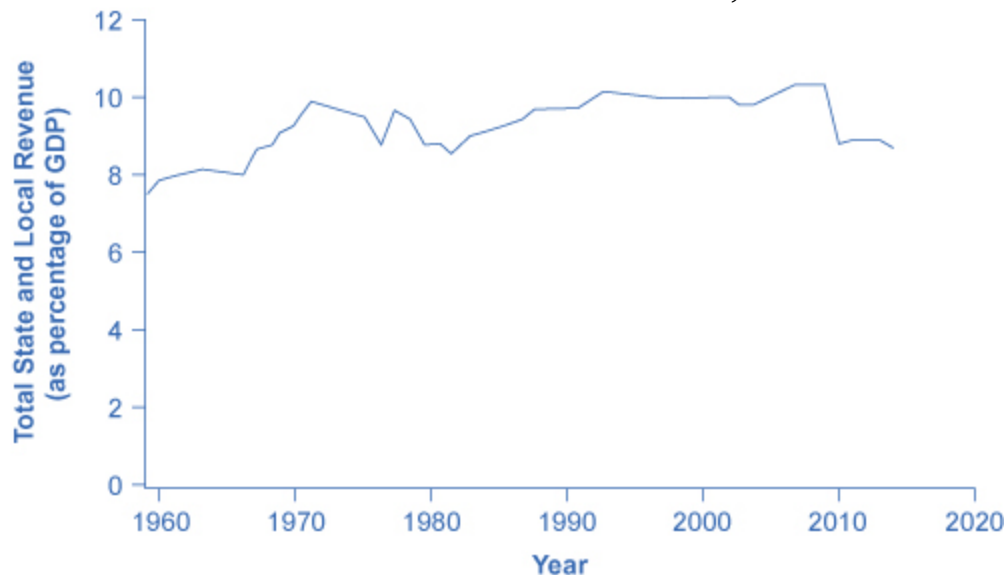
The federal government has a few other, smaller sources of revenue. It imposes an **excise tax**—that is, a tax on a particular good—on gasoline, tobacco, and alcohol. As a share of GDP, the amount collected by these taxes has stayed nearly constant over time, from about 2% of GDP in the 1960s to roughly 3% by 2014, according to the nonpartisan Congressional Budget Office. The government also imposes an **estate and gift tax** on people who pass large amounts of assets to the next generation—either after death or during life in the form of gifts. These estate and gift taxes collected about 0.2% of GDP in the first decade of the 2000s. By a quirk of legislation, the estate and gift tax was repealed in 2010, but reinstated in 2011. Other federal taxes, which are also relatively small in magnitude, include tariffs collected on imported goods and charges for inspections of goods entering the country.

State and Local Taxes

At the state and local level, taxes have been rising as a share of GDP over the last few decades to match the gradual rise in spending, as [\[link\]](#) illustrates. The main revenue sources for state and local governments are sales taxes, property taxes, and revenue passed along from the federal government, but many state and local governments also levy personal and corporate income taxes, as well as impose a wide variety of fees and

charges. The specific sources of tax revenue vary widely across state and local governments. Some states rely more on property taxes, some on sales taxes, some on income taxes, and some more on revenues from the federal government.

State and Local Tax Revenue as a Share of GDP, 1960–2014



State and local tax revenues have increased to match the rise in state and local spending. (Source: *Economic Report of the President, 2015*. Table B-21, <https://www.whitehouse.gov/administration/eop/cea/economic-report-of-the-President/2015>)

Key Concepts and Summary

The two main federal taxes are individual income taxes and payroll taxes that provide funds for Social Security and Medicare; these taxes together account for more than 80% of federal revenues. Other federal taxes include the corporate income tax, excise taxes on alcohol, gasoline and tobacco, and the estate and gift tax. A progressive tax is one, like the federal income tax, where those with higher incomes pay a higher share of taxes out of their income than those with lower incomes. A proportional tax is one, like the payroll tax for Medicare, where everyone pays the same share of taxes

regardless of income level. A regressive tax is one, like the payroll tax (above a certain threshold) that supports Social Security, where those with high income pay a lower share of income in taxes than those with lower incomes.

Self-Check Questions

Exercise:

Problem:

Suppose that gifts were taxed at a rate of 10% for amounts up to \$100,000 and 20% for anything over that amount. Would this tax be regressive or progressive?

Solution:

Progressive. People who give larger gifts subject to the higher tax rate would typically have larger incomes as well.

Exercise:

Problem:

If an individual owns a corporation for which he is the only employee, which different types of federal tax will he have to pay?

Solution:

Corporate income tax on his profits, individual income tax on his salary, and payroll tax taken out of the wages he pays himself.

Exercise:

Problem:

What taxes would an individual pay if he were self-employed and the business is not incorporated?

Solution:

individual income taxes

Exercise:

Problem:

The social security tax is 6.2% on employees' income earned below \$113,000. Is this tax progressive, regressive or proportional?

Solution:

The tax is regressive because wealthy income earners are not taxed at all on income above \$113,000. As a percent of total income, the social security tax hits lower income earners harder than wealthier individuals.

Review Questions

Exercise:

Problem:

What are the main categories of U.S. federal government taxes?

Exercise:

Problem:

What is the difference between a progressive tax, a proportional tax, and a regressive tax?

Critical Thinking Questions

Exercise:

Problem:

Excise taxes on tobacco and alcohol and state sales taxes are often criticized for being regressive. Although everyone pays the same rate regardless of income, why might this be so?

Exercise:

Problem:

What is the benefit of having state and local taxes on income instead of collecting all such taxes at the federal level?

References

Burman, Leonard E., and Joel Selmrod. *Taxes in America: What Everyone Needs to Know*. New York: Oxford University Press, 2012.

Hall, Robert E., and Alvin Rabushka. *The Flat Tax (Hoover Classics)*. Stanford: Hoover Institution Press, 2007.

Kliff, Sarah. “How Congress Paid for Obamacare (in Two Charts).” *The Washington Post: WonkBlog* (blog), August 30, 2012.
<http://www.washingtonpost.com/blogs/wonkblog/wp/2012/08/30/how-congress-paid-for-obamacare-in-two-charts/>.

Matthews, Dylan. “America’s Taxes are the Most Progressive in the World. Its Government is Among the Least.” *The Washington Post: WonkBlog* (blog). April 5, 2013.
<http://www.washingtonpost.com/blogs/wonkblog/wp/2013/04/05/americas-taxes-are-the-most-progressive-in-the-world-its-government-is-among-the-least/>.

Glossary

corporate income tax

a tax imposed on corporate profits

estate and gift tax

a tax on people who pass assets to the next generation—either after death or during life in the form of gifts

excise tax

a tax on a specific good—on gasoline, tobacco, and alcohol

individual income tax

a tax based on the income, of all forms, received by individuals

marginal tax rates

or the tax that must be paid on all yearly income

payroll tax

a tax based on the pay received from employers; the taxes provide funds for Social Security and Medicare

progressive tax

a tax that collects a greater share of income from those with high incomes than from those with lower incomes

proportional tax

a tax that is a flat percentage of income earned, regardless of level of income

regressive tax

a tax in which people with higher incomes pay a smaller share of their income in tax

Federal Deficits and the National Debt

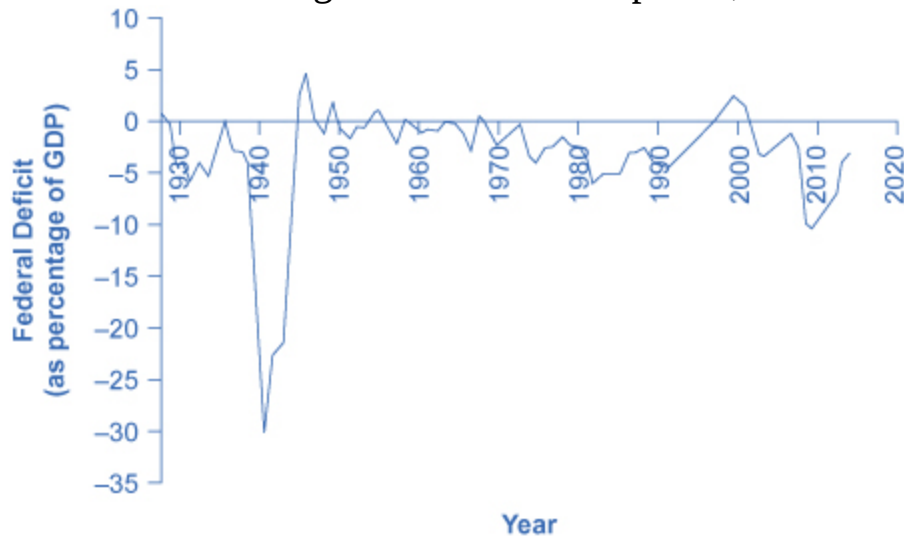
By the end of this section, you will be able to:

- Explain the U.S. federal budget in terms of annual debt and accumulated debt
- Understand how economic growth or decline can influence a budget surplus or budget deficit

Having discussed the revenue (taxes) and expense (spending) side of the budget, we now turn to the annual budget deficit or surplus, which is the difference between the tax revenue collected and spending over a fiscal year, which starts October 1 and ends September 30 of the next year.

[\[link\]](#) shows the pattern of annual federal budget deficits and surpluses, back to 1930, as a share of GDP. When the line is above the horizontal axis, the budget is in surplus; when the line is below the horizontal axis, a budget deficit occurred. Clearly, the biggest deficits as a share of GDP during this time were incurred to finance World War II. Deficits were also large during the 1930s, the 1980s, the early 1990s, and most recently during the recession of 2008–2009.

Pattern of Federal Budget Deficits and Surpluses, 1929–2014



The federal government has run budget deficits for decades. The budget was briefly in surplus in the late 1990s, before heading into deficit again in the first decade

of the 2000s—and especially deep deficits in the recession of 2008–2009. (Source: Federal Reserve Bank of St. Louis (FRED)).

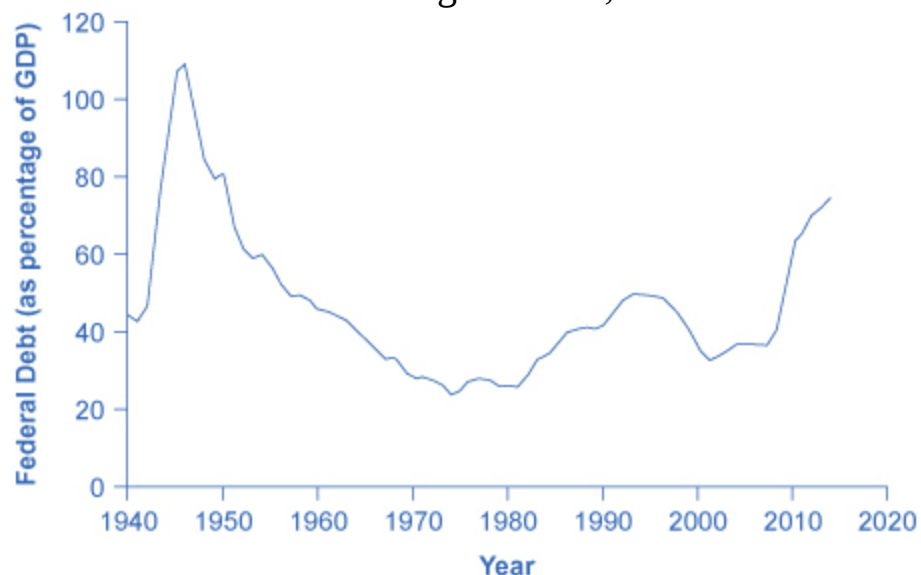
<http://research.stlouisfed.org/fred2/series/FYFSGDA188S>

)

Debt/GDP Ratio

Another useful way to view the budget deficit is through the prism of accumulated debt rather than annual deficits. The **national debt** refers to the total amount that the government has borrowed over time; in contrast, the budget deficit refers to how much has been borrowed in one particular year. [\[link\]](#) shows the ratio of debt/GDP since 1940. Until the 1970s, the debt/GDP ratio revealed a fairly clear pattern of federal borrowing. The government ran up large deficits and raised the debt/GDP ratio in World War II, but from the 1950s to the 1970s the government ran either surpluses or relatively small deficits, and so the debt/GDP ratio drifted down. Large deficits in the 1980s and early 1990s caused the ratio to rise sharply. When budget surpluses arrived from 1998 to 2001, the debt/GDP ratio declined substantially. The budget deficits starting in 2002 then tugged the debt/GDP ratio higher—with a big jump when the recession took hold in 2008–2009.

Federal Debt as a Percentage of GDP, 1942–2014



Federal debt is the sum of annual budget deficits and surpluses. Annual deficits do not always mean that the debt/GDP ratio is rising. During the 1960s and 1970s, the government often ran small deficits, but since the debt was growing more slowly than the economy, the debt/GDP ratio was declining over this time. In the 2008–2009 recession, the debt/GDP ratio rose sharply. (Source: *Economic Report of the President, Table B-20*, <http://www.gpo.gov/fdsys/pkg/ERP-2015/content-detail.html>)

The next Clear it Up feature discusses how the government handles the national debt.

Note:

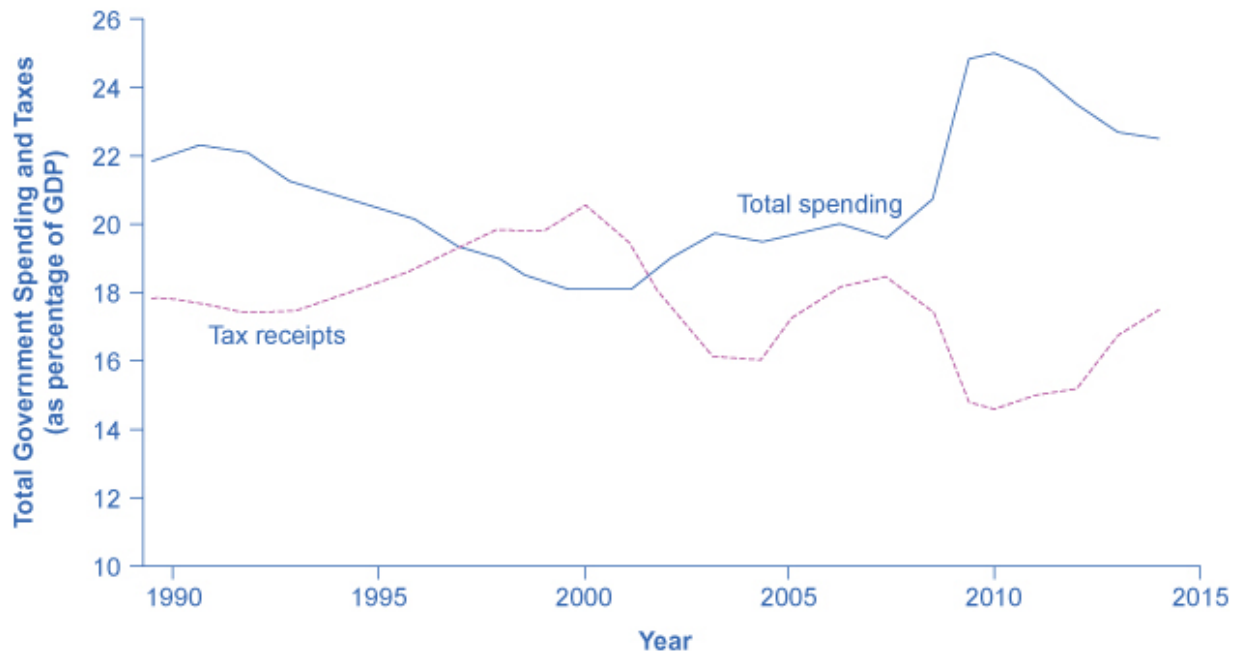
What is the national debt?

One year's federal budget deficit causes the federal government to sell Treasury bonds to make up the difference between spending programs and tax revenues. The dollar value of all the outstanding Treasury bonds on which the federal government owes money is equal to the national debt.

The Path from Deficits to Surpluses to Deficits

Why did the budget deficits suddenly turn to surpluses from 1998 to 2001? And why did the surpluses return to deficits in 2002? Why did the deficit become so large after 2007? [\[link\]](#) suggests some answers. The graph combines the earlier information on total federal spending and taxes in a single graph, but focuses on the federal budget since 1990.

Total Government Spending and Taxes as a Share of GDP, 1990–2014



When government spending exceeds taxes, the gap is the budget deficit. When taxes exceed spending, the gap is a budget surplus. The recessionary period starting in late 2007 saw higher spending and lower taxes, combining to create a large deficit in 2009. (Source: *Economic Report of the President, Tables B-21 and B-1*, "<http://www.gpo.gov/fdsys/pkg/ERP-2015/content-detail.html>")

Government spending as a share of GDP declined steadily through the 1990s. The biggest single reason was that defense spending declined from 5.2% of GDP in 1990 to 3.0% in 2000, but interest payments by the federal government also fell by about 1.0% of GDP. However, federal tax collections increased substantially in the later 1990s, jumping from 18.1% of GDP in 1994 to 20.8% in 2000. Powerful economic growth in the late 1990s fueled the boom in taxes. Personal income taxes rise as income goes up; payroll taxes rise as jobs and payrolls go up; corporate income taxes rise as profits go up. At the same time, government spending on transfer payments such as unemployment benefits, foods stamps, and welfare declined with more people working.

This sharp increase in tax revenues and decrease in expenditures on transfer payments was largely unexpected even by experienced budget analysts, and so budget surpluses came as a surprise. But in the early 2000s, many of these factors started running in reverse. Tax revenues sagged, due largely to the recession that started in March 2001, which reduced revenues. A series of tax cuts was enacted by Congress and signed into law by President George W. Bush, starting in 2001. In addition, government spending swelled due to increases in defense, healthcare, education, Social Security, and support programs for those who were hurt by the recession and the slow growth that followed. Deficits returned. When the severe recession hit in late 2007, spending climbed and tax collections fell to historically unusual levels, resulting in enormous deficits.

Longer-term forecasts of the U.S. budget, a decade or more into the future, predict enormous deficits. The higher deficits run during the recession of 2008–2009 have repercussions, and the demographics will be challenging. The primary reason is the “baby boom”—the exceptionally high birthrates that began in 1946, right after World War II, and lasted for about two decades. Starting in 2010, the front edge of the baby boom generation began to reach age 65, and in the next two decades, the proportion of Americans over the age of 65 will increase substantially. The current level of the payroll taxes that support Social Security and Medicare will fall well short of the projected expenses of these programs, as the following Clear It Up feature shows; thus, the forecast is for large budget deficits. A decision to collect more revenue to support these programs or to decrease benefit levels would alter this long-term forecast.

Note:

What is the long-term budget outlook for Social Security and Medicare? In 1946, just one American in 13 was over age 65. By 2000, it was one in eight. By 2030, one American in five will be over age 65. Two enormous U.S. federal programs focus on the elderly—Social Security and Medicare. The growing numbers of elderly Americans will increase spending on these programs, as well as on Medicaid. The current payroll tax levied on workers, which supports all of Social Security and the hospitalization

insurance part of Medicare, will not be enough to cover the expected costs. So, what are the options?

Long-term projections from the Congressional Budget Office in 2009 are that Medicare and Social Security spending combined will rise from 8.3% of GDP in 2009 to about 13% by 2035 and about 20% in 2080. If this rise in spending occurs, without any corresponding rise in tax collections, then some mix of changes must occur: (1) taxes will need to be increased dramatically; (2) other spending will need to be cut dramatically; (3) the retirement age and/or age receiving Medicare benefits will need to increase, or (4) the federal government will need to run extremely large budget deficits.

Some proposals suggest removing the cap on wages subject to the payroll tax, so that those with very high incomes would have to pay the tax on the entire amount of their wages. Other proposals suggest moving Social Security and Medicare from systems in which workers pay for retirees toward programs that set up accounts where workers save funds over their lifetimes and then draw out after retirement to pay for healthcare.

The United States is not alone in this problem. Indeed, providing the promised level of retirement and health benefits to a growing proportion of elderly with a falling proportion of workers is an even more severe problem in many European nations and in Japan. How to pay promised levels of benefits to the elderly will be a difficult public policy decision.

In the next module we shift to the use of fiscal policy to counteract business cycle fluctuations. In addition, we will explore proposals requiring a balanced budget—that is, for government spending and taxes to be equal each year. [The Impacts of Government Borrowing](#) will also cover how fiscal policy and government borrowing will affect national saving—and thus affect economic growth and trade imbalances.

Key Concepts and Summary

For most of the twentieth century, the U.S. government took on debt during wartime and then paid down that debt slowly in peacetime. However, it took on quite substantial debts in peacetime in the 1980s and early 1990s,

before a brief period of budget surpluses from 1998 to 2001, followed by a return to annual budget deficits since 2002, with very large deficits in the recession of 2008 and 2009. A budget deficit or budget surplus is measured annually. Total government debt or national debt is the sum of budget deficits and budget surpluses over time.

Self-Check Questions

Exercise:

Problem:

Debt has a certain self-reinforcing quality to it. There is one category of government spending that automatically increases along with the federal debt. What is it?

Solution:

As debt increases, interest payments also rise, so that the deficit grows even if we keep other government spending constant.

Exercise:

Problem: True or False:

- a. Federal spending has grown substantially in recent decades.
- b. By world standards, the U.S. government controls a relatively large share of the U.S. economy.
- c. A majority of the federal government's revenue is collected through personal income taxes.
- d. Education spending is slightly larger at the federal level than at the state and local level.
- e. State and local government spending has not risen much in recent decades.
- f. Defense spending is higher now than ever.
- g. The share of the economy going to federal taxes has increased substantially over time.

- h. Foreign aid is a large portion, although less than half, of federal spending.
 - i. Federal deficits have been very large for the last two decades.
 - j. The accumulated federal debt as a share of GDP is near an all-time high.
-

Solution:

- a. As a share of GDP, this is false. In nominal dollars, it is true.
- b. False.
- c. False.
- d. False. Education spending is much higher at the state level.
- e. False. As a share of GDP, it is up about 50.
- f. As a share of GDP, this is false, and in real dollars, it is also false.
- g. False.
- h. False; it's about 1%.
- i. False. Although budget deficits were large in 2003 and 2004, and continued into the later 2000s, the federal government ran budget surpluses from 1998–2001.
- j. False.

Review Questions

Exercise:

Problem:

What has been the general pattern of U.S. budget deficits in recent decades?

Exercise:

Problem:

What is the difference between a budget deficit and the national debt?

Critical Thinking Questions

Exercise:

Problem:

In a booming economy, is the federal government more likely to run surpluses or deficits? What are the various factors at play?

Exercise:

Problem:

Economist Arthur Laffer famously pointed out that, in some cases, income tax revenue can actually go up when tax rates go down. Why might this be the case?

Exercise:

Problem:

Is it possible for a nation to run budget deficits and still have its debt/GDP ratio fall? Explain your answer. Is it possible for a nation to run budget surpluses and still have its debt/GDP ratio rise? Explain your answer.

Problems

Exercise:

Problem:

If a government runs a budget deficit of \$10 billion dollars each year for ten years, then a surplus of \$1 billion for five years, and then a balanced budget for another ten years, what is the government debt?

References

Eisner, Robert. *The Great Deficit Scare: The Federal Budget, Trade, and Social Security*. New York: Priority Press Publications, 1997.

Weisman, Jonathan, and Ashley Parker. "Republicans Back Down, Ending Crisis Over Shutdown and Debt Limit." *The New York Times*, October 16, 2013. <http://www.nytimes.com/2013/10/17/us/congress-budget-debate.html>.

Wessel, David. *Red Ink: Inside the High-Stakes Politics of Federal Budget*. New York: Crown Publishing Group, 2013.

Glossary

national debt

the total accumulated amount the government has borrowed, over time, and not yet paid back

Using Fiscal Policy to Fight Recession, Unemployment, and Inflation

By the end of this section, you will be able to:

- Explain how expansionary fiscal policy can shift aggregate demand and influence the economy
- Explain how contractionary fiscal policy can shift aggregate demand and influence the economy

We need to emphasize that fiscal policy is the use of government spending and tax policy to alter the economy. Fiscal policy does not include all spending (such as the increase in spending that accompanies a war).

Aggregate demand and aggregate supply do not always move neatly together. Aggregate demand may fail to increase along with aggregate supply, or aggregate demand may even shift left, for a number of possible reasons: households become hesitant about consuming; firms decide against investing as much; or perhaps the demand from other countries for exports diminishes. For example, investment by private firms in physical capital in the U.S. economy boomed during the late 1990s, rising from 14.1% of GDP in 1993 to 17.2% in 2000, before falling back to 15.2% by 2002.

Conversely, if shifts in aggregate demand run ahead of increases in aggregate supply, inflationary increases in the price level will result. Business cycles of recession and recovery are the consequence of shifts in aggregate supply and aggregate demand.

[Monetary Policy and Bank Regulation](#) shows us that a central bank can use its powers over the banking system to engage in countercyclical—or “against the business cycle”—actions. If recession threatens, the central bank uses an expansionary monetary policy to increase the supply of money, increase the quantity of loans, reduce interest rates, and shift aggregate demand to the right. If inflation threatens, the central bank uses contractionary monetary policy to reduce the supply of money, reduce the quantity of loans, raise interest rates, and shift aggregate demand to the left. Fiscal policy is another macroeconomic policy tool for adjusting aggregate demand by using either government spending or taxation policy.

Expansionary Fiscal Policy

Expansionary fiscal policy increases the level of aggregate demand, through either increases in government spending or reductions in taxes.

Expansionary policy can do this by (1) increasing consumption by raising disposable income through cuts in personal income taxes or payroll taxes; (2) increasing investments by raising after-tax profits through cuts in business taxes; and (3) increasing government purchases through increased spending by the federal government on final goods and services and raising federal grants to state and local governments to increase their expenditures on final goods and services. Contractionary fiscal policy does the reverse: it decreases the level of aggregate demand by decreasing consumption, decreasing investments, and decreasing government spending, either through cuts in government spending or increases in taxes.

Should the government use tax cuts or spending increases, or a mix of the two, to carry out expansionary fiscal policy? After the Great Recession of 2008–2009 (which started, actually, in very late 2007), U.S. government spending rose from 19.6% of GDP in 2007 to 24.6% in 2009, while tax revenues declined from 18.5% of GDP in 2007 to 14.8% in 2009. The choice between whether to use tax or spending tools often has a political tinge. As a general statement, conservatives and Republicans prefer to see expansionary fiscal policy carried out by tax cuts, while liberals and Democrats prefer that expansionary fiscal policy be implemented through spending increases. The Obama administration and Congress passed an \$830 billion expansionary policy in early 2009 involving both tax cuts and increases in government spending, according to the Congressional Budget Office. However, state and local governments, whose budgets were also hard hit by the recession, began cutting their spending—a policy that offset federal expansionary policy.

The conflict over which policy tool to use can be frustrating to those who want to categorize economics as “liberal” or “conservative,” or who want to use economic models to argue against their political opponents. But the AD–AS model can be used both by advocates of smaller government, who seek to reduce taxes and government spending, and by advocates of bigger government, who seek to raise taxes and government spending. Economic

studies of specific taxing and spending programs can help to inform decisions about whether taxes or spending should be changed, and in what ways. Ultimately, decisions about whether to use tax or spending mechanisms to implement macroeconomic policy is, in part, a political decision rather than a purely economic one.

Contractionary Fiscal Policy

Fiscal policy can also contribute to pushing aggregate demand beyond potential GDP in a way that leads to inflation. A very large budget deficit pushes aggregate demand too far. This is sometimes known as an “overheating economy” where demand is so high that there is upward pressure on wages and prices, causing inflation. In this situation, contractionary fiscal policy involving federal spending cuts or tax increases can help to reduce the upward pressure on the price level by shifting aggregate demand to the left.

Again, the AD–AS model does not dictate how this contractionary fiscal policy is to be carried out. Some may prefer spending cuts; others may prefer tax increases; still others may say that it depends on the specific situation. The model only argues that, in this situation, aggregate demand needs to be reduced.

Key Concepts and Summary

Expansionary fiscal policy increases the level of aggregate demand, either through increases in government spending or through reductions in taxes. Expansionary fiscal policy is most appropriate when an economy is in recession and producing below its potential GDP. Contractionary fiscal policy decreases the level of aggregate demand, either through cuts in government spending or increases in taxes. Contractionary fiscal policy is most appropriate when an economy is producing above its potential GDP.

Self-Check Questions

Exercise:

Problem:

What is the main reason for employing contractionary fiscal policy in a time of strong economic growth?

Solution:

To keep prices from rising too much or too rapidly.

Exercise:**Problem:**

What is the main reason for employing expansionary fiscal policy during a recession?

Solution:

To increase employment.

Review Questions**Exercise:****Problem:**

What is the difference between expansionary fiscal policy and contractionary fiscal policy?

Exercise:**Problem:**

Under what general macroeconomic circumstances might a government use expansionary fiscal policy? When might it use contractionary fiscal policy?

Critical Thinking Questions

Exercise:**Problem:**

How will cuts in state budget spending affect federal expansionary policy?

Exercise:**Problem:**

Is expansionary fiscal policy more attractive to politicians who believe in larger government or to politicians who believe in smaller government? Explain your answer.

Problems**Exercise:****Problem:**

Specify whether expansionary or contractionary fiscal policy would seem to be most appropriate in response to each of the situations below and sketch a diagram using aggregate demand and aggregate supply curves to illustrate your answer:

- a. A recession.
- b. A stock market collapse that hurts consumer and business confidence.
- c. Extremely rapid growth of exports.
- d. Rising inflation.
- e. A rise in the natural rate of unemployment.
- f. A rise in oil prices.

References

Alesina, Alberto, and Francesco Giavazzi. *Fiscal Policy after the Financial Crisis (National Bureau of Economic Research Conference Report)*. Chicago: University Of Chicago Press, 2013.

Martin, Fernando M. “Fiscal Policy in the Great Recession and Lessons from the Past.” Federal Reserve Bank of St. Louis: *Economic Synopses*. no. 1 (2012). http://research.stlouisfed.org/publications/es/12/ES_2012-01-06.pdf.

Bivens, Josh, Andrew Fieldhouse, and Heidi Shierholz. “From Free-fall to Stagnation: Five Years After the Start of the Great Recession, Extraordinary Policy Measures Are Still Needed, But Are Not Forthcoming.” *Economic Policy Institute*. Last modified February 14, 2013. <http://www.epi.org/publication/bp355-five-years-after-start-of-great-recession/>.

Lucking, Brian, and Dan Wilson. Federal Reserve Bank of San Francisco, “FRBSF Economic Letter—U.S. Fiscal Policy: Headwind or Tailwind?” Last modified July 2, 2012. <http://www.frbsf.org/economic-research/publications/economic-letter/2012/july/us-fiscal-policy/>.

Greenstone, Michael, and Adam Looney. Brookings. “The Role of Fiscal Stimulus in the Ongoing Recovery.” Last modified July 6, 2012. <http://www.brookings.edu/blogs/jobs/posts/2012/07/06-jobs-greenstone-looney>.

Glossary

contractionary fiscal policy

fiscal policy that decreases the level of aggregate demand, either through cuts in government spending or increases in taxes

expansionary fiscal policy

fiscal policy that increases the level of aggregate demand, either through increases in government spending or cuts in taxes

The Multiplier Effect

The Multiplier Effect

Imagine an economy that operates at a GDP of \$800 when it's at a level of full employment, but the level of activity drops to \$700 in a short period of time. Government officials are urged to "do something" to revive the economy. By how much does government spending need to be increased so that the economy reaches the full employment GDP? The obvious answer might seem to be $\$800 - \$700 = \$100$; so raise government spending by \$100. But that answer is incorrect. A change of, for example, \$100 in government expenditures will have an effect of more than \$100 on the equilibrium level of real GDP. The reason is that a change in aggregate expenditures circles through the economy: households buy from firms, firms pay workers and suppliers, workers and suppliers buy goods from other firms, those firms pay their workers and suppliers, and so on. In this way, the original change in aggregate expenditures is actually spent more than once. This is called the multiplier effect: An initial increase in spending, cycles repeatedly through the economy and has a larger impact than the initial dollar amount spent.

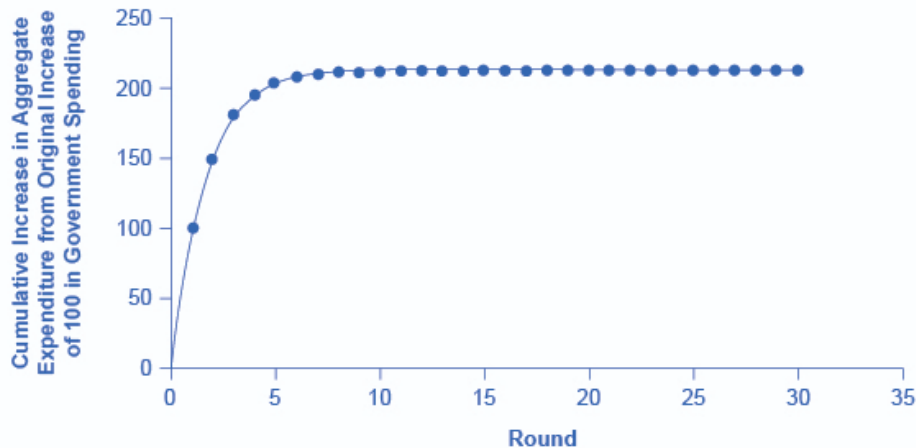
Side note: The following discussion relies on the Keynesian view that government has a role to play in some intervention during a recession. As we will learn later, this is not a view held by all schools of economic thought, but it has been widely practiced since the Great Depression.

How Does the Multiplier Work?

To understand how the multiplier effect works, return to the example in which the current equilibrium a real GDP of \$700, or \$100 short of the \$800 needed to be at full employment, potential GDP. If the government spends \$100 to close this gap, someone in the economy receives that spending and can treat it as income. Assume that those who receive this income pay 30% in taxes, save 10% of after-tax income, spend 10% of total income on imports, and then spend the rest on domestically produced goods and services.

As shown in the calculations in [\[link\]](#) and [\[link\]](#), out of the original \$100 in government spending, \$53 is left to spend on domestically produced goods and services. That \$53 which was spent, becomes income to someone, somewhere in the economy. Those who receive that income also pay 30% in taxes, save 10% of after-tax income, and spend 10% of total income on imports, as shown in [\[link\]](#), so that an additional \$28.09 (that is, $0.53 \times \$53$) is spent in the third round. The people who receive that income then pay taxes, save, and buy imports, and the amount spent in the fourth round is \$14.89 (that is, $0.53 \times \$28.09$).

The Multiplier Effect



An original increase of government spending of \$100 causes a rise in aggregate expenditure of \$100. But that \$100 is income to others in the economy, and after they save, pay taxes, and buy imports, they spend \$53 of that \$100 in a second round. In turn, that \$53 is income to others. Thus, the original government spending of \$100 is multiplied by these cycles of spending, but the impact of each successive cycle gets smaller and smaller. Given the numbers in this example, the original government spending increase of \$100 raises aggregate expenditure by \$213; therefore, the multiplier in this example is $\$213/\$100 = 2.13$.

Original increase in aggregate expenditure from government spending	100
Which is income to people throughout the economy: Pay 30% in taxes. Save 10% of after-tax income. Spend 10% of income on imports. Second-round increase of...	$70 - 7 - 10 = 53$
Which is \$53 of income to people through the economy: Pay 30% in taxes. Save 10% of after-tax income. Spend 10% of income on imports. Third-round increase of...	$37.1 - 3.71 - 5.3 = 28.09$
Which is \$28.09 of income to people through the economy: Pay 30% in taxes. Save 10% of after-tax income. Spend 10% of income on imports. Fourth-round increase of...	$19.663 - 1.9663 - 2.809 = 14.89$

Calculating the Multiplier Effect

Thus, over the first four rounds of aggregate expenditures, the impact of the original increase in government spending of \$100 creates a rise in aggregate expenditures of \$100 + \$53 + \$28.09 + \$14.89 = \$195.98. [\[link\]](#) shows these total aggregate expenditures after these first four rounds, and then the figure shows the total aggregate expenditures after 30 rounds. The additional boost to aggregate expenditures is shrinking in each round of consumption. After about 10 rounds, the additional increments are very small indeed—nearly invisible to the naked eye. After 30 rounds, the additional increments in each round are so small that they have no practical consequence. After 30 rounds, the cumulative value of the initial boost in aggregate expenditure is approximately \$213. Thus, the government spending increase of \$100 eventually, after many cycles, produced an increase of \$213 in aggregate expenditure and real GDP. In this example, the multiplier is $\$213/\$100 = 2.13$.

Calculating the Multiplier

Fortunately for everyone who is not carrying around a computer with a spreadsheet program to project the impact of an original increase in expenditures over 20, 50, or 100 rounds of spending, there is a formula for calculating the multiplier.

Equation:

$$\text{Spending Multiplier} = 1/(1 - \text{MPC} * (1 - \text{tax rate}) + \text{MPI})$$

The data from [\[link\]](#) and [\[link\]](#) is:

- Marginal Propensity to Save (MPS) = 30%
- Tax rate = 10%
- Marginal Propensity to Import (MPI) = 10%

The MPC is equal to $1 - \text{MPS}$, or 0.7. Therefore, the spending multiplier is:

Equation:

$$\begin{aligned}\text{Spending Multiplier} &= \frac{1}{1 - (0.7 - (0.10)(0.7) - 0.10)} \\ &= \frac{1}{0.47} \\ &= 2.13\end{aligned}$$

A change in spending of \$100 multiplied by the spending multiplier of 2.13 is equal to a change in GDP of \$213. Not coincidentally, this result is exactly what was calculated in [\[link\]](#) after many rounds of expenditures cycling through the economy.

The size of the multiplier is determined by what proportion of the marginal dollar of income goes into taxes, saving, and imports. These three factors are known as “leakages,” because they determine how much demand “leaks out” in each round of the multiplier effect. If the

leakages are relatively small, then each successive round of the multiplier effect will have larger amounts of demand, and the multiplier will be high. Conversely, if the leakages are relatively large, then any initial change in demand will diminish more quickly in the second, third, and later rounds, and the multiplier will be small. Changes in the size of the leakages—a change in the marginal propensity to save, the tax rate, or the marginal propensity to import—will change the size of the multiplier.

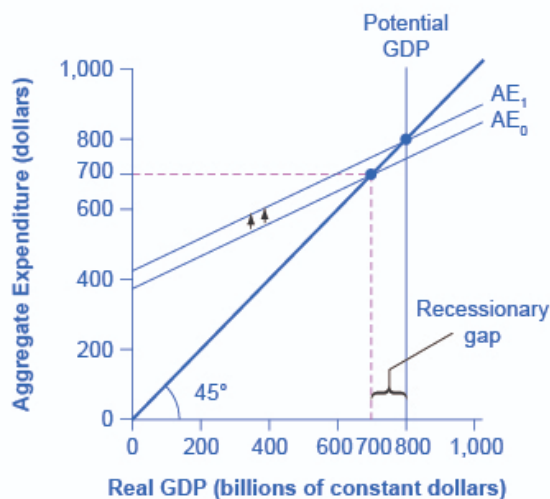
It may also help to remember that such increases in government spending are known as “injections” into the spending stream. Injections are generally increases in non-income determined spending; examples include increases in available credit, and increases in government spending.

When injections exceed leakages, the net effect is a positive multiplier, resulting in economic activity that is greater before the injections and leakages took place. When leakages are greater than injections, the result is that the level of economic activity has declined. Both are subject to a multiplier effect: the size of the impact is greater than the size of the injection (or leakage); how much bigger it is depends on the size of the multiplier.

Calculating Keynesian Policy Interventions

Returning to the original question: How much should government spending be increased to produce a total increase in real GDP of \$100? If the goal is to increase aggregate demand by \$100, and the multiplier is 2.13, then the increase in government spending to achieve that goal would be $\$100/2.13 = \47 . Government spending of approximately \$47, when combined with a multiplier of 2.13 (which is, remember, based on the specific assumptions about tax, saving, and import rates), produces an overall increase in real GDP of \$100, restoring the economy to potential GDP of \$800, as [\[link\]](#) shows.

The Multiplier Effect in an Expenditure-Output Model



The power of the multiplier effect is that an increase in expenditure has a larger increase on the equilibrium output. The

increase in expenditure is the vertical increase from AE_0 to AE_1 . However, the increase in equilibrium output, shown on the horizontal axis, is clearly larger.

The multiplier effect is also visible on the Keynesian cross diagram. [\[link\]](#) shows the example we have been discussing: a recessionary gap with an equilibrium of \$700, potential GDP of \$800, the slope of the aggregate expenditure function (AE_0) determined by the assumptions that taxes are 30% of income, savings are 0.1 of after-tax income, and imports are 0.1 of before-tax income. At AE_1 , the aggregate expenditure function is moved up to reach potential GDP.

Now, compare the vertical shift upward in the aggregate expenditure function, which is \$47, with the horizontal shift outward in real GDP, which is \$100 (as these numbers were calculated earlier). The rise in real GDP is more than double the rise in the aggregate expenditure function. (Similarly, if you look back at [\[link\]](#), you will see that the vertical movements in the aggregate expenditure functions are smaller than the change in equilibrium output that is produced on the horizontal axis. Again, this is the multiplier effect at work.) In this way, the power of the multiplier is apparent in the income–expenditure graph, as well as in the arithmetic calculation.

The multiplier does not just affect government spending, but applies to any change in the economy. Say that business confidence declines and investment falls off, or that the economy of a leading trading partner slows down so that export sales decline. These changes will reduce aggregate expenditures, and then will have an even larger effect on real GDP because of the multiplier effect. Read the following Clear It Up feature to learn how the multiplier effect can be applied to analyze the economic impact of professional sports.

Note:

How can the multiplier be used to analyze the economic impact of professional sports?

Attracting professional sports teams and building sports stadiums to create jobs and stimulate business growth is an economic development strategy adopted by many communities throughout the United States. In his recent article, “Public Financing of Private Sports Stadiums,” James Joyner of *Outside the Beltway* looked at public financing for NFL teams. Joyner’s findings confirm the earlier work of John Siegfried of Vanderbilt University and Andrew Zimbalist of Smith College.

Siegfried and Zimbalist used the multiplier to analyze this issue. They considered the amount of taxes paid and dollars spent locally to see if there was a positive multiplier effect. Since most professional athletes and owners of sports teams are rich enough to owe a lot of taxes, let’s say that 40% of any marginal income they earn is paid in taxes. Because athletes are often high earners with short careers, let’s assume that they save one-third of their after-tax income.

However, many professional athletes do not live year-round in the city in which they play, so let's say that one-half of the money that they do spend is spent outside the local area. One can think of spending outside a local economy, in this example, as the equivalent of imported goods for the national economy.

Now, consider the impact of money spent at local entertainment venues other than professional sports. While the owners of these other businesses may be comfortably middle-income, few of them are in the economic stratosphere of professional athletes. Because their incomes are lower, so are their taxes; say that they pay only 35% of their marginal income in taxes. They do not have the same ability, or need, to save as much as professional athletes, so let's assume their MPC is just 0.8. Finally, because more of them live locally, they will spend a higher proportion of their income on local goods—say, 65%.

If these general assumptions hold true, then money spent on professional sports will have less local economic impact than money spent on other forms of entertainment. For professional athletes, out of a dollar earned, 40 cents goes to taxes, leaving 60 cents. Of that 60 cents, one-third is saved, leaving 40 cents, and half is spent outside the area, leaving 20 cents. Only 20 cents of each dollar is cycled into the local economy in the first round. For locally-owned entertainment, out of a dollar earned, 35 cents goes to taxes, leaving 65 cents. Of the rest, 20% is saved, leaving 52 cents, and of that amount, 65% is spent in the local area, so that 33.8 cents of each dollar of income is recycled into the local economy.

Siegfried and Zimbalist make the plausible argument that, within their household budgets, people have a fixed amount to spend on entertainment. If this assumption holds true, then money spent attending professional sports events is money that was not spent on other entertainment options in a given metropolitan area. Since the multiplier is lower for professional sports than for other local entertainment options, the arrival of professional sports to a city would reallocate entertainment spending in a way that causes the local economy to shrink, rather than to grow. Thus, their findings seem to confirm what Joyner reports and what newspapers across the country are reporting. A quick Internet search for “economic impact of sports” will yield numerous reports questioning this economic development strategy.

Multiplier Tradeoffs: Stability versus the Power of Macroeconomic Policy

Is an economy healthier with a high multiplier or a low one? With a high multiplier, any change in aggregate demand will tend to be substantially magnified, and so the economy will be more unstable. With a low multiplier, by contrast, changes in aggregate demand will not be multiplied much, so the economy will tend to be more stable.

However, with a low multiplier, government policy changes in taxes or spending will tend to have less impact on the equilibrium level of real output. With a higher multiplier, government policies to raise or reduce aggregate expenditures will have a larger effect. Thus, a low multiplier means a more stable economy, but also weaker government macroeconomic policy, while a high multiplier means a more volatile economy, but also an economy in which government macroeconomic policy is more powerful.

Key Concepts and Summary

The expenditure-output model or Keynesian cross diagram shows how the level of aggregate expenditure (on the vertical axis) varies with the level of economic output (shown on the horizontal axis). Since the value of all macroeconomic output also represents income to someone somewhere else in the economy, the horizontal axis can also be interpreted as national income. The equilibrium in the diagram will occur where the aggregate expenditure line crosses the 45-degree line, which represents the set of points where aggregate expenditure in the economy is equal to output (or national income). Equilibrium in a Keynesian cross diagram can happen at potential GDP, or below or above that level.

The consumption function shows the upward-sloping relationship between national income and consumption. The marginal propensity to consume (MPC) is the amount consumed out of an additional dollar of income. A higher marginal propensity to consume means a steeper consumption function; a lower marginal propensity to consume means a flatter consumption function. The marginal propensity to save (MPS) is the amount saved out of an additional dollar of income. It is necessarily true that $MPC + MPS = 1$. The investment function is drawn as a flat line, showing that investment in the current year does not change with regard to the current level of national income. However, the investment function will move up and down based on the expected rate of return in the future. Government spending is drawn as a horizontal line in the Keynesian cross diagram, because its level is determined by political considerations, not by the current level of income in the economy. Taxes in the basic Keynesian cross diagram are taken into account by adjusting the consumption function. The export function is drawn as a horizontal line in the Keynesian cross diagram, because exports do not change as a result of changes in domestic income, but they move as a result of changes in foreign income, as well as changes in exchange rates. The import function is drawn as a downward-sloping line, because imports rise with national income, but imports are a subtraction from aggregate demand. Thus, a higher level of imports means a lower level of expenditure on domestic goods.

In a Keynesian cross diagram, the equilibrium may be at a level below potential GDP, which is called a recessionary gap, or at a level above potential GDP, which is called an inflationary gap.

The multiplier effect describes how an initial change in aggregate demand generated several times as much as cumulative GDP. The size of the spending multiplier is determined by three leakages: spending on savings, taxes, and imports. The formula for the multiplier is:

Equation:

$$\text{Multiplier} = \frac{1}{1 - (MPC \times (1 - \text{tax rate}) + MPI)}$$

An economy with a lower multiplier is more stable—it is less affected either by economic events or by government policy than an economy with a higher multiplier.

Self-Check Questions

Exercise:

Problem: An economy has the following characteristics:

Y = National income

Taxes = $T = 0.25Y$

C = Consumption = $400 + 0.85(Y - T)$

$I = 300$

$G = 200$

$X = 500$

$M = 0.1(Y - T)$

Find the equilibrium for this economy. If potential GDP is 3,500, then what change in government spending is needed to achieve this level? Do this problem two ways. First, plug 3,500 into the equations and solve for G . Second, calculate the multiplier and figure it out that way.

Solution

First, set up the calculation.

Equation:

$$AE = 400 + 0.85(Y - T) + 300 + 200 + 500 - 0.1(Y - T)$$

$$AE = Y$$

Then insert Y for AE and $0.25Y$ for T .

Equation:

$$Y = 400 + 0.85(Y - 0.25Y) + 300 + 200 + 500 - 0.1(Y - 0.25Y)$$

$$Y = 1400 + 0.6375Y - 0.075Y$$

$$0.4375Y = 1400$$

$$Y = 3200$$

If full employment is 3,500, then one approach is to plug in 3,500 for Y throughout the equation, but to leave G as a separate variable.

Equation:

$$\begin{aligned}
Y &= 400 + 0.85(Y - 0.25Y) + 300 + G + 500 + 0.1(Y - 0.25Y) \\
3500 &= 400 + 0.85(3500 - 0.25(3500)) + 300 + G + 500 - 0.1(3500 - 0.25(3500)) \\
G &= 3500 - 400 - 2231.25 - 1300 - 500 + 262.5 \\
G &= 331.25
\end{aligned}$$

A G value of 331.25 is an increase of 131.25 from its original level of 200.

Alternatively, the multiplier is that, out of every dollar spent, 0.25 goes to taxes, leaving 0.75, and out of after-tax income, 0.15 goes to savings and 0.1 to imports. Because $(0.75)(0.15) = 0.1125$ and $(0.75)(0.1) = 0.075$, this means that out of every dollar spent: $1 - 0.25 - 0.1125 - 0.075 = 0.5625$.

Thus, using the formula, the multiplier is:

Equation:

$$\frac{1}{1 - 0.5625} = 2.2837$$

To increase equilibrium GDP by 300, it will take a boost of $300/2.2837$, which again works out to 131.25.

Exercise:

Problem:

[\[link\]](#) represents the data behind a Keynesian cross diagram. Assume that the tax rate is 0.4 of national income; the MPC out of the after-tax income is 0.8; investment is \$2,000; government spending is \$1,000; exports are \$2,000 and imports are 0.05 of after-tax income. What is the equilibrium level of output for this economy?

National Income	After-tax Income	Consumption	I + G + X	Minus Imports	Aggregate Expenditures
\$8,000		\$4,340			
\$9,000					
\$10,000					

National Income	After- tax Income	Consumption	I + G + X	Minus Imports	Aggregate Expenditures
\$11,000					
\$12,000					
\$13,000					

Solution

The following table illustrates the completed table. The equilibrium level is italicized.

National Income	After- tax Income	Consumption	I + G + X	Minus Imports	Aggregate Expenditures
\$8,000	\$4,800	\$4,340	\$5,000	\$240	\$9,100
\$9,000	\$5,400	\$4,820	\$5,000	\$270	\$9,550
<i>\$10,000</i>	<i>\$6,000</i>	<i>\$5,300</i>	<i>\$5,000</i>	<i>\$300</i>	<i>\$10,000</i>
\$11,000	\$6,600	\$5,780	\$5,000	\$330	\$10,450
\$12,000	\$7,200	\$6,260	\$5,000	\$360	\$10,900
\$13,000	\$7,800	\$6,740	\$5,000	\$4,390	\$11,350

The alternative way of determining equilibrium is to solve for Y, where Y = national income, using: $Y = AE = C + I + G + X - M$

Equation:

$$Y = \$500 + 0.8(Y - T) + \$2,000 + \$1,000 + \$2,000 - 0.05(Y - T)$$

Solving for Y , we see that the equilibrium level of output is $Y = \$10,000$.

Exercise:

Problem: Explain how the multiplier works. Use an MPC of 80% in an example.

Solution

The multiplier refers to how many times a dollar will turnover in the economy. It is based on the Marginal Propensity to Consume (MPC) which tells how much of every dollar received will be spent. If the MPC is 80% then this means that out of every one dollar received by a consumer, \$0.80 will be spent. This \$0.80 is received by another person. In turn, 80% of the \$0.80 received, or \$0.64, will be spent, and so on. The impact of the multiplier is diluted when the effect of taxes and expenditure on imports is considered. To derive the multiplier, take the $1/1 - F$; where F is equal to percent of savings, taxes, and expenditures on imports.

Review Questions

Exercise:

Problem: What is on the axes of an expenditure-output diagram?

Exercise:

Problem: What does the 45-degree line show?

Exercise:

Problem: What determines the slope of a consumption function?

Exercise:

Problem:

What is the marginal propensity to consume, and how is it related to the marginal propensity to import?

Exercise:

Problem:

Why are the investment function, the government spending function, and the export function all drawn as flat lines?

Exercise:

Problem:

Why does the import function slope down? What is the marginal propensity to import?

Exercise:

Problem:

What are the components on which the aggregate expenditure function is based?

Exercise:

Problem:

Is the equilibrium in a Keynesian cross diagram usually expected to be at or near potential GDP?

Exercise:

Problem: What is an inflationary gap? A recessionary gap?

Exercise:

Problem: What is the multiplier effect?

Exercise:

Problem:

Why are savings, taxes, and imports referred to as “leakages” in calculating the multiplier effect?

Exercise:

Problem:

Will an economy with a high multiplier be more stable or less stable than an economy with a low multiplier in response to changes in the economy or in government policy?

Exercise:

Problem: How do economists use the multiplier?

Critical Thinking Questions

Exercise:

Problem:

What does it mean when the aggregate expenditure line crosses the 45-degree line? In other words, how would you explain the intersection in words?

Exercise:

Problem:

Which model, the AD/AS or the AE model better explains the relationship between rising price levels and GDP? Why?

Exercise:

Problem:

What are some reasons that the economy might be in a recession, and what is the appropriate government action to alleviate the recession?

Exercise:

Problem:

What should the government do to relieve inflationary pressures if the aggregate expenditure is greater than potential GDP?

Exercise:

Problem:

Two countries are in a recession. Country A has an MPC of 0.8 and Country B has an MPC of 0.6. In which country will government spending have the greatest impact?

Exercise:

Problem:

Compare two policies: a tax cut on income or an increase in government spending on roads and bridges. What are both the short-term and long-term impacts of such policies on the economy?

Exercise:

Problem:

What role does government play in stabilizing the economy and what are the tradeoffs that must be considered?

Exercise:

Problem:

If there is a recessionary gap of \$100 billion, should the government increase spending by \$100 billion to close the gap? Why? Why not?

Exercise:

Problem: What other changes in the economy can be evaluated by using the multiplier?

References

Joyner, James. Outside the Beltway. "Public Financing of Private Sports Stadiums." Last modified May 23, 2012. <http://www.outsidethebeltway.com/public-financing-of-private-sports-stadiums/>.

Siegfried, John J., and Andrew Zimbalist. "The Economics of Sports Facilities and Their Communities." *Journal of Economic Perspectives*. no. 3 (2000): 95-114. <http://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.14.3.95>.

Automatic Stabilizers

By the end of this section, you will be able to:

- Describe how discretionary fiscal policy can be used by the federal government to stabilize the economy.
- Identify examples of automatic stabilizers.
- Understand how a standardized employment budget can be used to identify automatic stabilizers.

The millions of unemployed in 2008–2009 could collect unemployment insurance benefits to replace some of their salaries. Federal fiscal policies include **discretionary fiscal policy**, when the government passes a new law that explicitly changes tax or spending levels. The stimulus package of 2009 is an example. Changes in tax and spending levels can also occur automatically, due to **automatic stabilizers**, such as unemployment insurance and food stamps, which are programs that are already laws that stimulate aggregate demand in a recession and hold down aggregate demand in a potentially inflationary boom.

Counterbalancing Recession and Boom

Consider first the situation where aggregate demand has risen sharply, causing the equilibrium to occur at a level of output above potential GDP. This situation will increase inflationary pressure in the economy. The policy prescription in this setting would be a dose of contractionary fiscal policy, implemented through some combination of higher taxes and lower spending. To some extent, *both* changes happen automatically. On the tax side, a rise in aggregate demand means that workers and firms throughout the economy earn more. Because taxes are based on personal income and corporate profits, a rise in aggregate demand automatically increases tax payments. On the spending side, stronger aggregate demand typically means lower unemployment and fewer layoffs, and so there is less need for government spending on unemployment benefits, welfare, Medicaid, and other programs in the social safety net.

The process works in reverse, too. If aggregate demand were to fall sharply so that a recession occurs, then the prescription would be for expansionary fiscal policy—some mix of tax cuts and spending increases. The lower level of aggregate demand and higher unemployment will tend to pull down personal incomes and corporate profits, an effect that will reduce the amount of taxes owed automatically. Higher unemployment and a weaker economy should lead to increased government spending on unemployment benefits, welfare, and other similar domestic programs. In 2009, the stimulus package included an extension in the time allowed to collect unemployment insurance. In addition, the automatic stabilizers react to a weakening of aggregate demand with expansionary fiscal policy and react to a strengthening of aggregate demand with contractionary fiscal policy, just as the AD/AS analysis suggests.

The very large budget deficit of 2009 was produced by a combination of automatic stabilizers and discretionary fiscal policy. The Great Recession, starting in late 2007, meant less tax-generating economic activity, which triggered the automatic stabilizers that reduce taxes. Most economists, even those who are concerned about a possible pattern of persistently large budget deficits, are much less concerned or even quite supportive of larger budget deficits in the short run of a few years during and immediately after a severe recession.

A glance back at economic history provides a second illustration of the power of automatic stabilizers. Remember that the length of economic upswings between recessions has become longer in the U.S. economy in recent decades (as discussed in [Unemployment](#)). The three longest economic booms of the twentieth century happened in the 1960s, the 1980s, and the 1991–2001 time period. One reason why the economy has tipped into recession less frequently in recent decades is that the size of government spending and taxes has increased in the second half of the twentieth century. Thus, the automatic stabilizing effects from spending and taxes are now larger than they were in the first half of the twentieth century. Around 1900, for example, federal spending was only about 2% of GDP. In 1929, just before the Great Depression hit, government spending was still just 4% of GDP. In those earlier times, the smaller size of government made

automatic stabilizers far less powerful than in the last few decades, when government spending often hovers at 20% of GDP or more.

The Standardized Employment Deficit or Surplus

Each year, the nonpartisan Congressional Budget Office (CBO) calculates the **standardized employment budget**—that is, what the budget deficit or surplus would be if the economy were producing at potential GDP, where people who look for work were finding jobs in a reasonable period of time and businesses were making normal profits, with the result that both workers and businesses would be earning more and paying more taxes. In effect, the standardized employment deficit eliminates the impact of the automatic stabilizers. [\[link\]](#) compares the actual budget deficits of recent decades with the CBO's standardized deficit.

Note: Visit this [website](#) to learn more from the Congressional Budget Office.

Comparison of Actual Budget Deficits with the Standardized Employment Deficit



When the economy is in recession, the standardized employment budget deficit is less than the actual budget deficit because the economy is below potential GDP, and the automatic stabilizers are reducing taxes and increasing spending. When the economy is performing extremely well, the standardized employment deficit (or surplus) is higher than the actual budget deficit (or surplus) because the economy is producing about potential GDP, so the automatic stabilizers are increasing taxes and reducing the need for government spending.

(Sources: *Actual and Cyclically Adjusted Budget Surpluses/Deficits*, <http://www.cbo.gov/publication/43977>; and *Economic Report of the President, Table B-1*, <http://www.gpo.gov/fdsys/pkg/ERP-2013/content-detail.html>)

Notice that in recession years, like the early 1990s, 2001, or 2009, the standardized employment deficit is smaller than the actual deficit. During recessions, the automatic stabilizers tend to increase the budget deficit, so if the economy was instead at full employment, the deficit would be reduced. However, in the late 1990s the standardized employment budget surplus was lower than the actual budget surplus. The gap between the standardized budget deficit or surplus and the actual budget deficit or surplus shows the impact of the automatic stabilizers. More generally, the standardized budget figures allow you to see what the budget deficit would look like with the economy held constant—at its potential GDP level of output.

Automatic stabilizers occur quickly. Lower wages means that a lower amount of taxes is withheld from paychecks right away. Higher unemployment or poverty means that government spending in those areas rises as quickly as people apply for benefits. However, while the automatic stabilizers offset part of the shifts in aggregate demand, they do not offset

all or even most of it. Historically, automatic stabilizers on the tax and spending side offset about 10% of any initial movement in the level of output. This offset may not seem enormous, but it is still useful. Automatic stabilizers, like shock absorbers in a car, can be useful if they reduce the impact of the worst bumps, even if they do not eliminate the bumps altogether.

Key Concepts and Summary

Fiscal policy is conducted both through discretionary fiscal policy, which occurs when the government enacts taxation or spending changes in response to economic events, or through automatic stabilizers, which are taxing and spending mechanisms that, by their design, shift in response to economic events without any further legislation. The standardized employment budget is the calculation of what the budget deficit or budget surplus would have been in a given year if the economy had been producing at its potential GDP in that year. Many economists and politicians criticize the use of fiscal policy for a variety of reasons, including concerns over time lags, the impact on interest rates, and the inherently political nature of fiscal policy. We cover the critique of fiscal policy in the next module.

Self-Check Questions

Exercise:

Problem:

In a recession, does the actual budget surplus or deficit fall above or below the standardized employment budget?

Solution:

It falls below because less tax revenue than expected is collected.

Exercise:

Problem:

What is the main advantage of automatic stabilizers over discretionary fiscal policy?

Solution:

Automatic stabilizers take effect very quickly, whereas discretionary policy can take a long time to implement.

Exercise:**Problem:**

Explain how automatic stabilizers work, both on the taxation side and on the spending side, first in a situation where the economy is producing less than potential GDP and then in a situation where the economy is producing more than potential GDP.

Solution:

In a recession, because of the decline in economic output, less income is earned, and so less in taxes is automatically collected. Many welfare and unemployment programs are designed so that those who fall into certain categories, like “unemployed” or “low income,” are eligible for benefits. During a recession, more people fall into these categories and become eligible for benefits automatically. The combination of reduced taxes and higher spending is just what is needed for an economy in recession producing below potential GDP. With an economic boom, average income levels rise in the economy, so more in taxes is automatically collected. Fewer people meet the criteria for receiving government assistance to the unemployed or the needy, so government spending on unemployment assistance and welfare falls automatically. This combination of higher taxes and lower spending is just what is needed if an economy is producing above its potential GDP.

Review Questions

Exercise:

Problem:

What is the difference between discretionary fiscal policy and automatic stabilizers?

Exercise:

Problem: Why do automatic stabilizers function “automatically?”

Exercise:

Problem: What is the standardized employment budget?

Critical Thinking Questions

Exercise:

Problem:

Is Medicaid (federal government aid to low-income families and individuals) an automatic stabilizer?

Glossary

automatic stabilizers

tax and spending rules that have the effect of slowing down the rate of decrease in aggregate demand when the economy slows down and restraining aggregate demand when the economy speeds up, without any additional change in legislation

discretionary fiscal policy

the government passes a new law that explicitly changes overall tax or spending levels with the intent of influencing the level or overall economic activity

standardized employment budget

the budget deficit or surplus in any given year adjusted for what it would have been if the economy were producing at potential GDP

Practical Problems with Discretionary Fiscal Policy

By the end of this section, you will be able to:

- Understand how fiscal policy and monetary policy are interconnected
- Explain the three lag times that often occur when solving economic problems.
- Identify the legal and political challenges of responding to an economic problem.

In the early 1960s, many leading economists believed that the problem of the business cycle, and the swings between cyclical unemployment and inflation, were a thing of the past. On the cover of its December 31, 1965, issue, *Time* magazine, then the premier news magazine in the United States, ran a picture of John Maynard Keynes, and the story inside identified Keynesian theories as “the prime influence on the world’s economies.” The article reported that policymakers have “used Keynesian principles not only to avoid the violent [business] cycles of prewar days but to produce phenomenal economic growth and to achieve remarkably stable prices.”

This happy consensus, however, did not last. The U.S. economy suffered one recession from December 1969 to November 1970, a deeper recession from November 1973 to March 1975, and then double-dip recessions from January to June 1980 and from July 1981 to November 1982. At various times, inflation and unemployment both soared. Clearly, the problems of macroeconomic policy had not been completely solved. As economists began to consider what had gone wrong, they identified a number of issues that make discretionary fiscal policy more difficult than it had seemed in the rosy optimism of the mid-1960s.

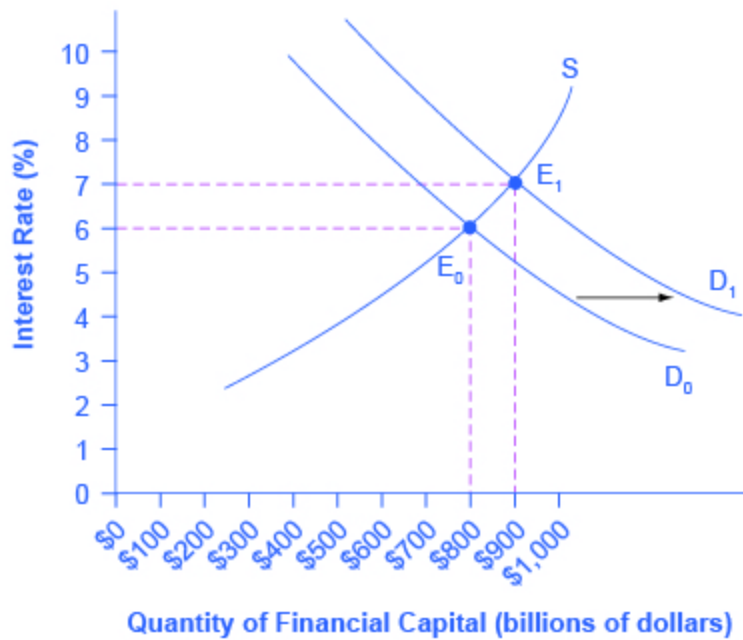
Fiscal Policy and Interest Rates

Because fiscal policy affects the quantity that the government borrows in financial capital markets, it not only affects aggregate demand—it can also affect interest rates. In [\[link\]](#), the original equilibrium (E_0) in the financial capital market occurs at a quantity of \$800 billion and an interest rate of 6%. However, an increase in government budget deficits shifts the demand

for financial capital from D_0 to D_1 . The new equilibrium (E_1) occurs at a quantity of \$900 billion and an interest rate of 7%.

A consensus estimate based on a number of studies is that an increase in budget deficits (or a fall in budget surplus) by 1% of GDP will cause an increase of 0.5–1.0% in the long-term interest rate.

Fiscal Policy and Interest Rates



When a government borrows money in the financial capital market, it causes a shift in the demand for financial capital from D_0 to D_1 . As the equilibrium moves from E_0 to E_1 , the equilibrium interest rate rises from 6% to 7% in this example. In this way, an expansionary fiscal policy intended to shift aggregate demand to the right can also lead to a higher interest rate, which has the effect of shifting aggregate demand back to the left.

A problem arises here. An expansionary fiscal policy, with tax cuts or spending increases, is intended to increase aggregate demand. If an expansionary fiscal policy also causes higher interest rates, then firms and households are discouraged from borrowing and spending (as occurs with tight monetary policy), thus reducing aggregate demand. Even if the direct effect of expansionary fiscal policy on increasing demand is not totally offset by lower aggregate demand from higher interest rates, fiscal policy can end up being less powerful than was originally expected. This is referred to as **crowding out**, where government borrowing and spending results in higher interest rates, which reduces business investment and household consumption.

The broader lesson is that fiscal and monetary policy must be coordinated. If expansionary fiscal policy is to work well, then the central bank can also reduce or keep short-term interest rates low. Conversely, monetary policy can also help to ensure that contractionary fiscal policy does not lead to a recession.

Long and Variable Time Lags

Monetary policy can be changed several times each year, but fiscal policy is much slower to be enacted. Imagine that the economy starts to slow down. It often takes some months before the economic statistics signal clearly that a downturn has started, and a few months more to confirm that it is truly a recession and not just a one- or two-month blip. The time it takes to determine that a recession has occurred is often called the **recognition lag**. After this lag, policymakers become aware of the problem and propose fiscal policy bills. The bills go into various congressional committees for hearings, negotiations, votes, and then, if passed, eventually for the president's signature. Many fiscal policy bills about spending or taxes propose changes that would start in the next budget year or would be phased in gradually over time. The time to get a bill passed is often referred to as the **legislative lag**. Finally, once the bill is passed it takes some time for the funds to be dispersed to the appropriate agencies to implement the programs. The time to get the projects started is often called the **implementation lag**.

Moreover, the exact level of fiscal policy to be implemented is never completely clear. Should the budget deficit be increased by 0.5% of GDP? By 1% of GDP? By 2% of GDP? In an AD/AS diagram, it is straightforward to sketch an aggregate demand curve shifting to the potential GDP level of output. In the real world, the actual level of potential output is known only roughly, not precisely, and exactly how a spending cut or tax increase will affect aggregate demand is always somewhat controversial. Also unknown is the state of the economy at any point in time. During the early days of the Obama administration, for example, no one knew how deep in the hole the economy really was. During the financial crisis of 2008-09, the rapid collapse of the banking system and automotive sector made it difficult to assess how quickly the economy was collapsing.

Thus, it can take many months or even more than a year to begin an expansionary fiscal policy after a recession has started—and even then, uncertainty will remain over exactly how much to expand or contract taxes and spending. When politicians attempt to use countercyclical fiscal policy to fight recession or inflation, they run the risk of responding to the macroeconomic situation of two or three years ago, in a way that may be exactly wrong for the economy at that time. George P. Schultz, a professor of economics, former Secretary of the Treasury, and Director of the Office of Management and Budget, once wrote: “While the economist is accustomed to the concept of lags, the politician likes instant results. The tension comes because, as I have seen on many occasions, the economist’s lag is the politician’s nightmare.”

Temporary and Permanent Fiscal Policy

A temporary tax cut or spending increase will explicitly last only for a year or two, and then revert back to its original level. A permanent tax cut or spending increase is expected to stay in place for the foreseeable future. The effect of temporary and permanent fiscal policies on aggregate demand can be very different. Consider how you would react if the government announced a tax cut that would last one year and then be repealed, in comparison with how you would react if the government announced a

permanent tax cut. Most people and firms will react more strongly to a permanent policy change than a temporary one.

This fact creates an unavoidable difficulty for countercyclical fiscal policy. The appropriate policy may be to have an expansionary fiscal policy with large budget deficits during a recession, and then a contractionary fiscal policy with budget surpluses when the economy is growing well. But if both policies are explicitly temporary ones, they will have a less powerful effect than a permanent policy.

Structural Economic Change Takes Time

When an economy recovers from a recession, it does not usually revert back to its exact earlier shape. Instead, the internal structure of the economy evolves and changes and this process can take time. For example, much of the economic growth of the mid-2000s was in the sectors of construction (especially of housing) and finance. However, when housing prices started falling in 2007 and the resulting financial crunch led into recession (as discussed in [Monetary Policy and Bank Regulation](#)), both sectors contracted. The manufacturing sector of the U.S. economy has been losing jobs in recent years as well, under pressure from technological change and foreign competition. Many of the people thrown out of work from these sectors in the Great Recession of 2008–2009 will never return to the same jobs in the same sectors of the economy; instead, the economy will need to grow in new and different directions, as the following Clear It Up feature shows. Fiscal policy can increase overall demand, but the process of structural economic change—the expansion of a new set of industries and the movement of workers to those industries—inevitably takes time.

Note:

Why do jobs vanish?

People can lose jobs for a variety of reasons: because of a recession, but also because of longer-run changes in the economy, such as new technology. Productivity improvements in auto manufacturing, for example, can reduce the number of workers needed, and eliminate these

jobs in the long run. The Internet has created jobs but also caused the loss of jobs as well, from travel agents to book store clerks. Many of these jobs may never come back. Short-run fiscal policy to reduce unemployment can create jobs, but it cannot replace jobs that will never return.

The Limitations of Fiscal Policy

Fiscal policy can help an economy that is producing below its potential GDP to expand aggregate demand so that it produces closer to potential GDP, thus lowering unemployment. But fiscal policy cannot help an economy produce at an output level above potential GDP without causing inflation. At this point, unemployment becomes so low that workers become scarce and wages rise rapidly.

Political Realities and Discretionary Fiscal Policy

A final problem for discretionary fiscal policy arises out of the difficulties of explaining to politicians how countercyclical fiscal policy that runs against the tide of the business cycle should work. Politicians often have a gut-level belief that when the economy and tax revenues slow down, it is time to hunker down, pinch pennies, and trim expenses. Countercyclical policy, however, says that when the economy has slowed down, it is time for the government to go on a spree, raising spending, and cutting taxes. This offsets the drop in the economy in the other sectors. Conversely, when economic times are good and tax revenues are rolling in, politicians often feel that it is time for tax cuts and new spending. But countercyclical policy says that this economic boom should be an appropriate time for keeping taxes high and restraining spending.

Politicians tend to prefer expansionary fiscal policy over contractionary policy. There is rarely a shortage of proposals for tax cuts and spending increases, especially during recessions. However, politicians are less willing to hear the message that in good economic times, they should propose tax increases and spending limits. In the economic upswing of the late 1990s and early 2000s, for example, the U.S. GDP grew rapidly. Estimates from

respected government economic forecasters like the nonpartisan Congressional Budget Office and the Office of Management and Budget stated that the GDP was above potential GDP, and that unemployment rates were unsustainably low. However, no mainstream politician took the lead in saying that the booming economic times might be an appropriate time for spending cuts or tax increases.

Discretionary Fiscal Policy: Summing Up

Expansionary fiscal policy can help to end recessions and contractionary fiscal policy can help to reduce inflation. Given the uncertainties over interest rate effects, time lags, temporary and permanent policies, and unpredictable political behavior, many economists and knowledgeable policymakers had concluded by the mid-1990s that discretionary fiscal policy was a blunt instrument, more like a club than a scalpel. It might still make sense to use it in extreme economic situations, like an especially deep or long recession. For less extreme situations, it was often preferable to let fiscal policy work through the automatic stabilizers and focus on monetary policy to steer short-term countercyclical efforts.

Key Concepts and Summary

Because fiscal policy affects the quantity of money that the government borrows in financial capital markets, it not only affects aggregate demand—it can also affect interest rates. If an expansionary fiscal policy also causes higher interest rates, then firms and households are discouraged from borrowing and spending, reducing aggregate demand in a situation called crowding out. Given the uncertainties over interest rate effects, time lags (implementation lag, legislative lag, and recognition lag), temporary and permanent policies, and unpredictable political behavior, many economists and knowledgeable policymakers have concluded that discretionary fiscal policy is a blunt instrument and better used only in extreme situations.

Self-Check Questions

Exercise:

Problem:

What would happen if expansionary fiscal policy was implemented in a recession but, due to lag, did not actually take effect until after the economy was back to potential GDP?

Solution:

Prices would be pushed up as a result of too much spending.

Exercise:**Problem:**

What would happen if contractionary fiscal policy were implemented during an economic boom but, due to lag, it did not take effect until the economy slipped into recession?

Solution:

Employment would suffer as a result of too little spending.

Exercise:**Problem:**

Do you think the typical time lag for fiscal policy is likely to be longer or shorter than the time lag for monetary policy? Explain your answer?

Solution:

Monetary policy probably has shorter time lags than fiscal policy. Imagine that the data becomes fairly clear that an economy is in or near a recession. Expansionary monetary policy can be carried out through open market operations, which can be done fairly quickly, since the Federal Reserve's Open Market Committee meets six times a year. Also, monetary policy takes effect through interest rates, which can change fairly quickly. However, fiscal policy is carried out through acts of Congress that need to be signed into law by the president. Negotiating such laws often takes months, and even after the laws are

negotiated, it takes more months for spending programs or tax cuts to have an effect on the macroeconomy.

Review Questions

Exercise:

Problem:

What are some practical weaknesses of discretionary fiscal policy?

Critical Thinking Questions

Exercise:

Problem:

What is a potential problem with a temporary tax increase designed to increase aggregate demand if people know that it is temporary?

Exercise:

Problem:

If the government gives a \$300 tax cut to everyone in the country, explain the mechanism by which this will cause interest rates to rise.

References

Leduc, Sylvain, and Daniel Wilson. Federal Reserve Bank of San Francisco: Working Paper Series. "Are State Governments Roadblocks to Federal Stimulus? Evidence from Highway Grants in the 2009 Recovery Act. (Working Paper 2013-16)." Last modified July 2013. <http://www.frbsf.org/economic-research/files/wp2013-16.pdf>.

Lucking, Brian, and Daniel Wilson. "FRBSF Economic Letter-Fiscal Headwinds: Is the Other Shoe About to Drop?" *Federal Reserve Bank of San Francisco*. Last modified June 3, 2013. <http://www.frbsf.org/economic->

research/publications/economic-letter/2013/june/fiscal-headwinds-federal-budget-policy/.

Recovery.gov. “Track the Money.”
<http://www.recovery.gov/Pages/default.aspx>.

Bastagli, Francesca, David Coady, and Sanjeev Gupta. International Monetary Fund. “IMF Staff Discussion Note: Income Inequality and Fiscal Policy.” Last modified June 28, 2012.
<http://www.imf.org/external/pubs/ft/sdn/2012/sdn1208.pdf>.

Glossary

crowding out

federal spending and borrowing causes interest rates to rise and business investment to fall

implementation lag

the time it takes for the funds relating to fiscal policy to be dispersed to the appropriate agencies to implement the programs

legislative lag

the time it takes to get a fiscal policy bill passed

recognition lag

the time it takes to determine that a recession has occurred

The Question of a Balanced Budget

By the end of this section, you will be able to:

- Understand the arguments for and against requiring the U.S. federal budget to be balanced
- Consider the long-run and short-run effects of a federal budget deficit

For many decades, going back to the 1930s, proposals have been put forward to require that the U.S. government balance its budget every year. In 1995, a proposed constitutional amendment that would require a balanced budget passed the U.S. House of Representatives by a wide margin, and failed in the U.S. Senate by only a single vote. (For the balanced budget to have become an amendment to the Constitution would have required a two-thirds vote by Congress and passage by three-quarters of the state legislatures.)

Most economists view the proposals for a perpetually balanced budget with bemusement. After all, in the short term, economists would expect the budget deficits and surpluses to fluctuate up and down with the economy and the automatic stabilizers. Economic recessions should automatically lead to larger budget deficits or smaller budget surpluses, while economic booms lead to smaller deficits or larger surpluses. A requirement that the budget be balanced each and every year would prevent these automatic stabilizers from working and would worsen the severity of economic fluctuations.

Some supporters of the balanced budget amendment like to argue that, since households must balance their own budgets, the government should too. But this analogy between household and government behavior is severely flawed. Most households do not balance their budgets every year. Some years households borrow to buy houses or cars or to pay for medical expenses or college tuition. Other years they repay loans and save funds in retirement accounts. After retirement, they withdraw and spend those savings. Also, the government is not a household for many reasons, one of which is that the government has macroeconomic responsibilities. The argument of Keynesian macroeconomic policy is that the government needs

to lean against the wind, spending when times are hard and saving when times are good, for the sake of the overall economy.

There is also no particular reason to expect a government budget to be balanced in the medium term of a few years. For example, a government may decide that by running large budget deficits, it can make crucial long-term investments in human capital and physical infrastructure that will build the long-term productivity of a country. These decisions may work out well or poorly, but they are not always irrational. Such policies of ongoing government budget deficits may persist for decades. As the U.S. experience from the end of World War II up to about 1980 shows, it is perfectly possible to run budget deficits almost every year for decades, but as long as the percentage increases in debt are smaller than the percentage growth of GDP, the debt/GDP ratio will decline at the same time.

Nothing in this argument should be taken as a claim that budget deficits are always a wise policy. In the short run, a government that runs a very large budget deficit can shift aggregate demand to the right and trigger severe inflation. Additionally, governments may borrow for foolish or impractical reasons. [The Macroeconomic Impacts of Government Borrowing](#) will discuss how large budget deficits, by reducing national saving, can in certain cases reduce economic growth and even contribute to international financial crises. A requirement that the budget be balanced in each calendar year, however, is a misguided overreaction to the fear that in some cases, budget deficits can become too large.

Note:

No Yellowstone Park?

The federal budget shutdown of 2013 illustrated the many sides to fiscal policy and the federal budget. In 2013, Republicans and Democrats could not agree on which spending policies to fund and how large the government debt should be. Due to the severity of the recession in 2008–2009, the fiscal stimulus, and previous policies, the federal budget deficit and debt was historically high. One way to try to cut federal spending and borrowing was to refuse to raise the legal federal debt limit, or tie on conditions to appropriation bills to stop the Affordable Health Care Act.

This disagreement led to a two-week shutdown of the federal government and got close to the deadline where the federal government would default on its Treasury bonds. Finally, however, a compromise emerged and default was avoided. This shows clearly how closely fiscal policies are tied to politics.

Key Concepts and Summary

Balanced budget amendments are a popular political idea, but the economic merits behind such proposals are questionable. Most economists accept that fiscal policy needs to be flexible enough to accommodate unforeseen expenditures, such as wars or recessions. While persistent, large budget deficits can indeed be a problem, a balanced budget amendment prevents even small, temporary deficits that might, in some cases, be necessary.

Self-Check Questions

Exercise:

Problem:

How would a balanced budget amendment affect a decision by Congress to grant a tax cut during a recession?

Solution:

The government would have to make up the revenue either by raising taxes in a different area or cutting spending.

Exercise:

Problem:

How would a balanced budget amendment change the effect of automatic stabilizer programs?

Solution:

Programs where the amount of spending is not fixed, but rather determined by macroeconomic conditions, such as food stamps, would lose a great deal of flexibility if spending increases had to be met by corresponding tax increases or spending cuts.

Review Questions

Exercise:

Problem:

What are some of the arguments for and against a requirement that the federal government budget be balanced every year?

Critical Thinking Questions

Exercise:

Problem:

Do you agree or disagree with this statement: “It is in the best interest of our economy for Congress and the President to run a balanced budget each year.” Explain your answer.

Exercise:

Problem:

During the Great Recession of 2008–2009, what actions would have been required of Congress and the President had a balanced budget amendment to the Constitution been ratified? What impact would that have had on the unemployment rate?

Introduction to Money and Banking
class="introduction"
Cowrie Shell or Money?

Is this an image
of a cowrie
shell or money?

The answer is:

Both. For
centuries, the
extremely
durable cowrie
shell was used
as a medium of
exchange in
various parts of
the world.

(Credit:
modification of
work by
“prilfish”/Flick
r Creative
Commons)



Note:

The Many Disguises of Money: From Cowries to Bitcoins

Here is a trivia question: In the history of the world, what item was used for money over the broadest geographic area and for the longest period of time? The answer is not gold, silver, or any precious metal. It is the cowrie, a mollusk shell found mainly off the Maldives Islands in the Indian Ocean. Cowries served as money as early as 700 B.C. in China. By the 1500s, they were in widespread use across India and Africa. For several centuries after that, cowries were used in markets including southern Europe, western Africa, India, and China for a wide range of purchases: everything from buying lunch or a ferry ride to paying for a shipload of silk or rice. Cowries were still acceptable as a way of paying taxes in certain African nations in the early twentieth century.

What made cowries work so well as money? First, they are extremely durable—lasting a century or more. As the late economic historian Karl Polanyi put it, they can be “poured, sacked, shoveled, hoarded in heaps” while remaining “clean, dainty, stainless, polished, and milk-white.”

Second, parties could use cowries either by counting shells of a certain

size, or—for large purchases—by measuring the weight or volume of the total shells to be exchanged. Third, it was impossible to counterfeit a cowrie shell, but gold or silver coins could be counterfeited by making copies with cheaper metals. Finally, in the heyday of cowrie money, from the 1500s into the 1800s, the collection of cowries was tightly controlled, first by the Portuguese and later by the Dutch and the English. As a result, the supply of cowries was allowed to grow quickly enough to serve the needs of commerce, but not so quickly that they were no longer scarce. Money throughout the ages has taken many different forms and continues to evolve even today. What do you think money is?

Note:

Introduction to Money and Banking

In this chapter, you will learn about:

- Defining Money by Its Functions
- Measuring Money: Currency, M1, and M2
- The Role of Banks
- How Banks Create Money

The discussion of money and banking is a central component in the study of macroeconomics. At this point, you should have firmly in mind the main goals of macroeconomics from [Welcome to Economics!](#): economic growth, low unemployment, and low inflation. We have yet to discuss money and its role in helping to achieve our macroeconomic goals.

You should also understand Keynesian and neoclassical frameworks for macroeconomic analysis and how these frameworks can be embodied in the aggregate demand/aggregate supply (AD/AS) model. With the goals and frameworks for macroeconomic analysis in mind, the final step is to discuss the two main categories of macroeconomic policy: monetary policy, which focuses on money, banking and interest rates; and fiscal policy, which focuses on government spending, taxes, and borrowing. This chapter

discusses what economists mean by money, and how money is closely interrelated with the banking system. [Monetary Policy and Bank Regulation](#) furthers this discussion.

Defining Money by Its Functions

By the end of this section, you will be able to:

- Explain the various functions of money
- Contrast commodity money and fiat money

Money for the sake of money is not an end in itself. You cannot eat dollar bills or wear your bank account. Ultimately, the usefulness of money rests in exchanging it for goods or services. As the American writer and humorist Ambrose Bierce (1842–1914) wrote in 1911, money is a “blessing that is of no advantage to us excepting when we part with it.” Money is what people regularly use when purchasing or selling goods and services, and thus money must be widely accepted by both buyers and sellers. This concept of money is intentionally flexible, because money has taken a wide variety of forms in different cultures.

Barter and the Double Coincidence of Wants

To understand the usefulness of money, we must consider what the world would be like without money. How would people exchange goods and services? Economies without money typically engage in the barter system. **Barter**—literally trading one good or service for another—is highly inefficient for trying to coordinate the trades in a modern advanced economy. In an economy without money, an exchange between two people would involve a **double coincidence of wants**, a situation in which two people each want some good or service that the other person can provide. For example, if an accountant wants a pair of shoes, this accountant must find someone who has a pair of shoes in the correct size and who is willing to exchange the shoes for some hours of accounting services. Such a trade is likely to be difficult to arrange. Think about the complexity of such trades in a modern economy, with its extensive division of labor that involves thousands upon thousands of different jobs and goods.

Another problem with the barter system is that it does not allow us to easily enter into future contracts for the purchase of many goods and services. For example, if the goods are perishable it may be difficult to exchange them

for other goods in the future. Imagine a farmer wanting to buy a tractor in six months using a fresh crop of strawberries. Additionally, while the barter system might work adequately in small economies, it will keep these economies from growing. The time that individuals would otherwise spend producing goods and services and enjoying leisure time is spent bartering.

Functions for Money

Money solves the problems created by the barter system. (We will get to its definition soon.) First, money serves as a **medium of exchange**, which means that money acts as an intermediary between the buyer and the seller. Instead of exchanging accounting services for shoes, the accountant now exchanges accounting services for money. This money is then used to buy shoes. To serve as a medium of exchange, money must be very widely accepted as a method of payment in the markets for goods, labor, and financial capital.

Second, money must serve as a **store of value**. In a barter system, we saw the example of the shoemaker trading shoes for accounting services. But she risks having her shoes go out of style, especially if she keeps them in a warehouse for future use—their value will decrease with each season. Shoes are not a good store of value. Holding money is a much easier way of storing value. You know that you do not need to spend it immediately because it will still hold its value the next day, or the next year. This function of money does not require that money is a *perfect* store of value. In an economy with inflation, money loses some buying power each year, but it remains money.

Third, money serves as a **unit of account**, which means that it is the ruler by which other values are measured. For example, an accountant may charge \$100 to file your tax return. That \$100 can purchase two pair of shoes at \$50 a pair. Money acts as a common denominator, an accounting method that simplifies thinking about trade-offs.

Finally, another function of money is that money must serve as a **standard of deferred payment**. This means that if money is usable today to make purchases, it must also be acceptable to make purchases today that will be

paid in the *future*. Loans and future agreements are stated in monetary terms and the standard of deferred payment is what allows us to buy goods and services today and pay in the future. So **money** serves all of these functions — it is a medium of exchange, store of value, unit of account, and standard of deferred payment.

Commodity versus Fiat Money

Money has taken a wide variety of forms in different cultures. Gold, silver, cowrie shells, cigarettes, and even cocoa beans have been used as money. Although these items are used as **commodity money**, they also have a value from use as something other than money. Gold, for example, has been used throughout the ages as money although today it is not used as money but rather is valued for its other attributes. Gold is a good conductor of electricity and is used in the electronics and aerospace industry. Gold is also used in the manufacturing of energy efficient reflective glass for skyscrapers and is used in the medical industry as well. Of course, gold also has value because of its beauty and malleability in the creation of jewelry.

As commodity money, gold has historically served its purpose as a medium of exchange, a store of value, and as a unit of account. **Commodity-backed currencies** are dollar bills or other currencies with values backed up by gold or other commodity held at a bank. During much of its history, the money supply in the United States was backed by gold and silver.

Interestingly, antique dollars dated as late as 1957, have “Silver Certificate” printed over the portrait of George Washington, as shown in [\[link\]](#). This meant that the holder could take the bill to the appropriate bank and exchange it for a dollar’s worth of silver.

A Silver Certificate and a Modern U.S. Bill



Until 1958, silver certificates were commodity-backed money—backed by silver, as indicated by the words “Silver Certificate” printed on the bill. Today, U.S. bills are backed by the Federal Reserve, but as fiat money. (Credit: “The.Comedian”/Flickr Creative Commons)

As economies grew and became more global in nature, the use of commodity monies became more cumbersome. Countries moved towards the use of **fiat money**. Fiat money has no intrinsic value, but is declared by a government to be the legal tender of a country. The United States’ paper money, for example, carries the statement: “THIS NOTE IS LEGAL TENDER FOR ALL DEBTS, PUBLIC AND PRIVATE.” In other words, by government decree, if you owe a debt, then legally speaking, you can pay that debt with the U.S. currency, even though it is not backed by a commodity. The only backing of our money is universal faith and trust that the currency has value, and nothing more.

Note: Watch this [video](#) on the “History of Money.”

Key Concepts and Summary

Money is what people in a society regularly use when purchasing or selling goods and services. If money were not available, people would need to barter with each other, meaning that each person would need to identify others with whom they have a double coincidence of wants—that is, each party has a specific good or service that the other desires. Money serves several functions: a medium of exchange, a unit of account, a store of value, and a standard of deferred payment. There are two types of money: commodity money, which is an item used as money, but which also has value from its use as something other than money; and fiat money, which has no intrinsic value, but is declared by a government to be the legal tender of a country.

Self-Check Questions

Exercise:

Problem:

In many casinos, a person buys chips to use for gambling. Within the walls of the casino, these chips can often be used to buy food and drink or even a hotel room. Do chips in a gambling casino serve all three functions of money?

Solution:

As long as you remain within the walls of the casino, chips fit the definition of money; that is, they serve as a medium of exchange, a unit of account, and a store of value. Chips do not work very well as money once you leave the casino, but many kinds of money do not work well in other areas. For example, it is hard to spend money from Turkey or Brazil at your local supermarket or at the movie theater.

Exercise:**Problem:**

Can you name some item that is a store of value, but does not serve the other functions of money?

Solution:

Many physical items that a person buys at one time but may sell at another time can serve as an answer to this question. Examples include a house, land, art, rare coins or stamps, and so on.

Review Questions**Exercise:**

Problem: What are the four functions served by money?

Exercise:**Problem:**

How does the existence of money simplify the process of buying and selling?

Exercise:

Problem: What is the double-coincidence of wants?

Critical Thinking Questions**Exercise:**

Problem:

The Bring it Home Feature discusses the use of cowrie shells as money. Although cowrie shells are no longer used as money, do you think other forms of commodity monies are possible? What role might technology play in our definition of money?

Exercise:**Problem:**

Imagine that you are a barber in a world without money. Explain why it would be tricky to obtain groceries, clothing, and a place to live.

References

Hogendorn, Jan and Marion Johnson. *The Shell Money of the Slave Trade*. Cambridge University Press, 2003. 6.

Glossary

barter

literally, trading one good or service for another, without using money

commodity money

an item that is used as money, but which also has value from its use as something other than money

commodity-backed currencies

are dollar bills or other currencies with values backed up by gold or another commodity

double coincidence of wants

a situation in which two people each want some good or service that the other person can provide

fiat money

has no intrinsic value, but is declared by a government to be the legal tender of a country

medium of exchange

whatever is widely accepted as a method of payment

money

whatever serves society in four functions: as a medium of exchange, a store of value, a unit of account, and a standard of deferred payment.

standard of deferred payment

money must also be acceptable to make purchases today that will be paid in the future

store of value

something that serves as a way of preserving economic value that can be spent or consumed in the future

unit of account

the common way in which market values are measured in an economy

Measuring Money: Currency, M1, and M2

By the end of this section, you will be able to:

- Contrast M1 money supply and M2 money supply
- Classify monies as M1 money supply or M2 money supply

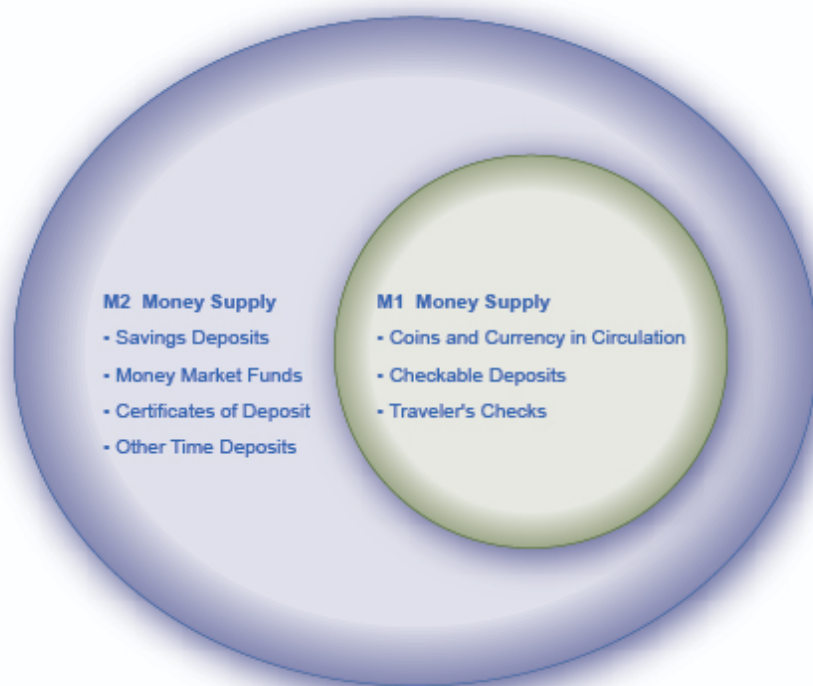
Cash in your pocket certainly serves as money. But what about checks or credit cards? Are they money, too? Rather than trying to state a single way of measuring money, economists offer broader definitions of money based on liquidity. Liquidity refers to how quickly a financial asset can be used to buy a good or service. For example, cash is very liquid. Your \$10 bill can be easily used to buy a hamburger at lunchtime. However, \$10 that you have in your savings account is not so easy to use. You must go to the bank or ATM machine and withdraw that cash to buy your lunch. Thus, \$10 in your savings account is *less* liquid.

The Federal Reserve Bank, which is the central bank of the United States, is a bank regulator and is responsible for monetary policy and defines money according to its liquidity. There are two definitions of money: M1 and M2 money supply. **M1 money supply** includes those monies that are very liquid such as cash, checkable (demand) deposits, and traveler's checks. **M2 money supply** is less liquid in nature and includes M1 plus savings and time deposits, certificates of deposits, and money market funds.

M1 money supply includes **coins and currency in circulation**—the coins and bills that circulate in an economy that are not held by the U.S. Treasury, at the Federal Reserve Bank, or in bank vaults. Closely related to currency are checkable deposits, also known as **demand deposits**. These are the amounts held in checking accounts. They are called demand deposits or checkable deposits because the banking institution must give the deposit holder his money “on demand” when a check is written or a debit card is used. These items together—currency, and checking accounts in banks—make up the definition of money known as M1, which is measured daily by the Federal Reserve System. Traveler's checks are also included in M1, but have decreased in use over the recent past.

A broader definition of money, M2 includes everything in M1 but also adds other types of deposits. For example, M2 includes **savings deposits** in banks, which are bank accounts on which you cannot write a check directly, but from which you can easily withdraw the money at an automatic teller machine or bank. Many banks and other financial institutions also offer a chance to invest in **money market funds**, where the deposits of many individual investors are pooled together and invested in a safe way, such as short-term government bonds. Another ingredient of M2 are the relatively small (that is, less than about \$100,000) certificates of deposit (CDs) or **time deposits**, which are accounts that the depositor has committed to leaving in the bank for a certain period of time, ranging from a few months to a few years, in exchange for a higher interest rate. In short, all these types of M2 are money that you can withdraw and spend, but which require a greater effort to do so than the items in M1 [\[link\]](#) should help in visualizing the relationship between M1 and M2. Note that M1 is included in the M2 calculation.

The Relationship between M1 and M2 Money



M1 and M2 money have several definitions, ranging from narrow to broad. M1 = coins and currency in circulation + checkable (demand) deposit + traveler's

checks. $M2 = M1 + \text{savings deposits} + \text{money market funds} + \text{certificates of deposit} + \text{other time deposits}$.

The Federal Reserve System is responsible for tracking the amounts of M1 and M2 and prepares a weekly release of information about the money supply. To provide an idea of what these amounts sound like, according to the Federal Reserve Bank's measure of the U.S. money stock, at the end of February 2015, M1 in the United States was \$3 trillion, while M2 was \$11.8 trillion. A breakdown of the portion of each type of money that comprised M1 and M2 in February 2015, as provided by the Federal Reserve Bank, is provided in [\[link\]](#).

Components of M1 in the U.S. (February 2015, Seasonally Adjusted)	\$ billions
Currency	\$1,271.8
Traveler's checks	\$2.9
Demand deposits and other checking accounts	\$1,713.5
<i>Total M1</i>	<i>\$2,988.2 (or \$3 trillion)</i>
Components of M2 in the U.S. (February 2015, Seasonally Adjusted)	\$ billions
M1 money supply	\$2,988.2

Savings accounts	\$7,712.1
Time deposits	\$509.2
Individual money market mutual fund balances	\$610.8
<i>Total M2</i>	<i>\$11,820.3 (or \$11.8 trillion)</i>

M1 and M2 Federal Reserve Statistical Release, Money Stock Measures (Source: Federal Reserve Statistical Release, <http://www.federalreserve.gov/RELEASES/h6/current/default.htm#t2tg1link>)

The lines separating M1 and M2 can become a little blurry. Sometimes elements of M1 are not treated alike; for example, some businesses will not accept personal checks for large amounts, but will accept traveler's checks or cash. Changes in banking practices and technology have made the savings accounts in M2 more similar to the checking accounts in M1. For example, some savings accounts will allow depositors to write checks, use automatic teller machines, and pay bills over the Internet, which has made it easier to access savings accounts. As with many other economic terms and statistics, the important point is to know the strengths and limitations of the various definitions of money, not to believe that such definitions are as clear-cut to economists as, say, the definition of nitrogen is to chemists.

Where does "plastic money" like debit cards, credit cards, and smart money fit into this picture? A **debit card**, like a check, is an instruction to the user's bank to transfer money directly and immediately from your bank account to the seller. It is important to note that in our definition of money, it is *checkable deposits* that are money, not the paper check or the debit card. Although you can make a purchase with a **credit card**, it is not considered money but rather a short term loan from the credit card company to you. When you make a purchase with a credit card, the credit card company immediately transfers money from its checking account to the seller, and at the end of the month, the credit card company sends you a bill for what you have charged that month. Until you pay the credit card bill, you have effectively borrowed money from the credit card company. With a **smart**

card, you can store a certain value of money on the card and then use the card to make purchases. Some “smart cards” used for specific purposes, like long-distance phone calls or making purchases at a campus bookstore and cafeteria, are not really all that smart, because they can only be used for certain purchases or in certain places.

In short, credit cards, debit cards, and smart cards are different ways to move money when a purchase is made. But having more credit cards or debit cards does not change the quantity of money in the economy, any more than having more checks printed increases the amount of money in your checking account.

One key message underlying this discussion of M1 and M2 is that money in a modern economy is not just paper bills and coins; instead, money is closely linked to bank accounts. Indeed, the macroeconomic policies concerning money are largely conducted through the banking system. The next section explains how banks function and how a nation’s banking system has the power to create money.

Key Concepts and Summary

Money is measured with several definitions: M1 includes currency and money in checking accounts (demand deposits). Traveler’s checks are also a component of M1, but are declining in use. M2 includes all of M1, plus savings deposits, time deposits like certificates of deposit, and money market funds.

Self-Check Questions

Exercise:

Problem:

If you are out shopping for clothes and books, what is easiest and most convenient for you to spend: M1 or M2? Explain your answer.

Solution:

The currency and checks in M1 are easiest to spend. It is harder to spend M2 directly, although if there is an automatic teller machine in the shopping mall, you can turn M2 from your savings account into an M1 of currency quite quickly. If your answer is about “credit cards,” then you are really talking about spending M1—although it is M1 from the account of the credit card company, which you will repay later when your credit card bill comes due.

Exercise:

Problem:

For the following list of items, indicate if they are in M1, M2, or neither:

- a. Your \$5,000 line of credit on your Bank of America card
- b. \$50 dollars’ worth of traveler’s checks you have not used yet
- c. \$1 in quarters in your pocket
- d. \$1200 in your checking account
- e. \$2000 you have in a money market account

Solution:

- a. Neither in M1 or M2
- b. That is part of M1, and because M2 includes M1 it is also part of M2
- c. Currency out in the public hands is part of M1 and M2
- d. Checking deposits are in M1 and M2
- e. Money market accounts are in M2

Review Questions

Exercise:

Problem: What components of money are counted as part of M1?

Exercise:

Problem: What components of money are counted in M2?

Critical Thinking Questions

Exercise:

Problem:

Explain why you think the Federal Reserve Bank tracks M1 and M2.

Exercise:

Problem:

The total amount of U.S. currency in circulation divided by the U.S. population comes out to about \$3,500 per person. That is more than most of us carry. Where is all the cash?

Exercise:

Problem:

If you take \$100 out of your piggy bank and deposit it in your checking account, how did M1 change? Did M2 change?

References

Federal Reserve Statistical Release. November 23, 2013.

<http://www.federalreserve.gov/RELEASES/h6/current/default.htm#t2tg1link>
.

Glossary

coins and currency in circulation

the coins and bills that circulate in an economy that are not held by the U.S Treasury, at the Federal Reserve Bank, or in bank vaults

credit card

immediately transfers money from the credit card company's checking account to the seller, and at the end of the month the user owes the money to the credit card company; a credit card is a short-term loan

debit card

like a check, is an instruction to the user's bank to transfer money directly and immediately from your bank account to the seller

demand deposit

checkable deposit in banks that is available by making a cash withdrawal or writing a check

M1 money supply

a narrow definition of the money supply that includes currency and checking accounts in banks, and to a lesser degree, traveler's checks.

M2 money supply

a definition of the money supply that includes everything in M1, but also adds savings deposits, money market funds, and certificates of deposit

money market fund

the deposits of many investors are pooled together and invested in a safe way like short-term government bonds

savings deposit

bank account where you cannot withdraw money by writing a check, but can withdraw the money at a bank—or can transfer it easily to a checking account

smart card

stores a certain value of money on a card and then the card can be used to make purchases

time deposit

account that the depositor has committed to leaving in the bank for a certain period of time, in exchange for a higher rate of interest; also

called certificate of deposit

The Role of Banks

By the end of this section, you will be able to:

- Explain how banks act as intermediaries between savers and borrowers
- Evaluate the relationship between banks, savings and loans, and credit unions
- Analyze the causes of bankruptcy and recessions

The late bank robber named Willie Sutton was once asked why he robbed banks. He answered: “That’s where the money is.” While this may have been true at one time, from the perspective of modern economists, Sutton is both right and wrong. He is wrong because the overwhelming majority of money in the economy is not in the form of currency sitting in vaults or drawers at banks, waiting for a robber to appear. Most money is in the form of bank accounts, which exist only as electronic records on computers. From a broader perspective, however, the bank robber was more right than he may have known. Banking is intimately interconnected with money and consequently, with the broader economy.

Banks make it far easier for a complex economy to carry out the extraordinary range of transactions that occur in goods, labor, and financial capital markets. Imagine for a moment what the economy would be like if all payments had to be made in cash. When shopping for a large purchase or going on vacation you might need to carry hundreds of dollars in a pocket or purse. Even small businesses would need stockpiles of cash to pay workers and to purchase supplies. A bank allows people and businesses to store this money in either a checking account or savings account, for example, and then withdraw this money as needed through the use of a direct withdrawal, writing a check, or using a debit card.

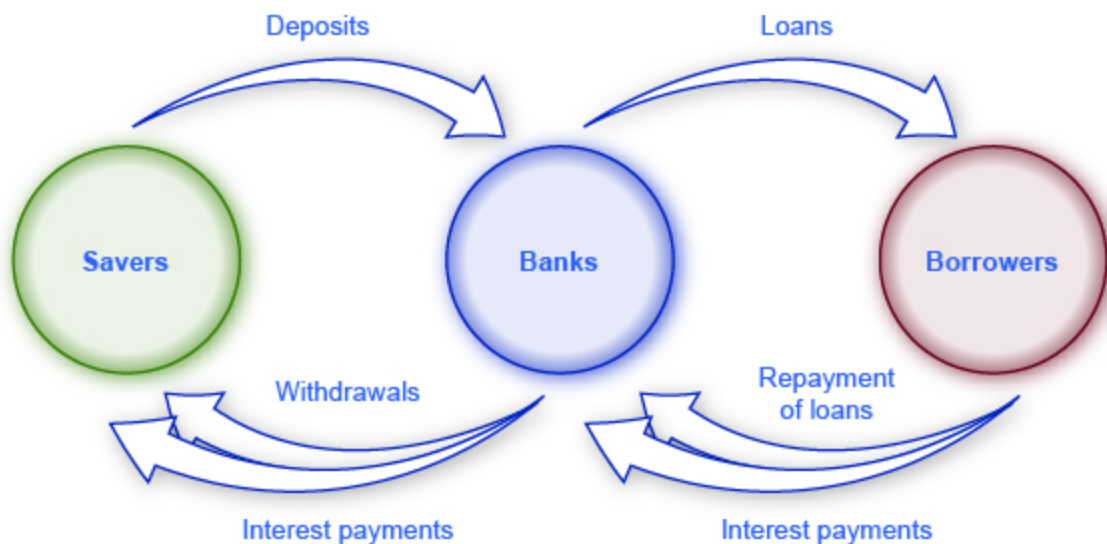
Banks are a critical intermediary in what is called the **payment system**, which helps an economy exchange goods and services for money or other financial assets. Also, those with extra money that they would like to save can store their money in a bank rather than look for an individual that is willing to borrow it from them and then repay them at a later date. Those who want to borrow money can go directly to a bank rather than trying to find someone to lend them cash **Transaction costs** are the costs associated

with finding a lender or a borrower for this money. Thus, banks lower transactions costs and act as financial intermediaries—they bring savers and borrowers together. Along with making transactions much safer and easier, banks also play a key role in the creation of money.

Banks as Financial Intermediaries

An “intermediary” is one who stands between two other parties. Banks are a **financial intermediary**—that is, an institution that operates between a saver who deposits money in a bank and a borrower who receives a loan from that bank. Financial intermediaries include other institutions in the financial market such as insurance companies and pension funds, but they will not be included in this discussion because they are not considered to be **depository institutions**, which are institutions that accept money *deposits* and then use these to make loans. All the funds deposited are mingled in one big pool, which is then loaned out. [\[link\]](#) illustrates the position of banks as financial intermediaries, with deposits flowing into a bank and loans flowing out. Of course, when banks make loans to firms, the banks will try to funnel financial capital to healthy businesses that have good prospects for repaying the loans, not to firms that are suffering losses and may be unable to repay.

Banks as Financial Intermediaries



Banks act as financial intermediaries because they stand between

savers and borrowers. Savers place deposits with banks, and then receive interest payments and withdraw money. Borrowers receive loans from banks and repay the loans with interest. In turn, banks return money to savers in the form of withdrawals, which also include interest payments from banks to savers.

Note:

How are banks, savings and loans, and credit unions related?

Banks have a couple of close cousins: savings institutions and credit unions. Banks, as explained, receive deposits from individuals and businesses and make loans with the money. Savings institutions are also sometimes called “savings and loans” or “thrifts.” They also take loans and make deposits. However, from the 1930s until the 1980s, federal law limited how much interest savings institutions were allowed to pay to depositors. They were also required to make most of their loans in the form of housing-related loans, either to homebuyers or to real-estate developers and builders.

A credit union is a nonprofit financial institution that its members own and run. Members of each credit union decide who is eligible to be a member. Usually, potential members would be everyone in a certain community, or groups of employees, or members of a certain organization. The credit union accepts deposits from members and focuses on making loans back to its members. While there are more credit unions than banks and more banks than savings and loans, the total assets of credit unions are growing. In 2008, there were 7,085 banks. Due to the bank failures of 2007–2009 and bank mergers, there were 5,571 banks in the United States at the end of the fourth quarter in 2014. According to the Credit Union National Association, as of December 2014 there were 6,535 credit unions with assets totaling \$1.1 billion. A day of “Transfer Your Money” took place in 2009 out of general public disgust with big bank bailouts. People were encouraged to transfer their deposits to credit unions. This has grown into the ongoing Move Your Money Project. Consequently, some now hold deposits as large as \$50 billion. However, as of 2013, the 12 largest banks

(0.2%) controlled 69 percent of all banking assets, according to the Dallas Federal Reserve.

A Bank's Balance Sheet

A **balance sheet** is an accounting tool that lists assets and liabilities. An **asset** is something of value that is owned and can be used to produce something. For example, the cash you own can be used to pay your tuition. If you own a home, this is also considered an asset. A **liability** is a debt or something you owe. Many people borrow money to buy homes. In this case, a home is the asset, but the mortgage is the liability. The **net worth** is the asset value minus how much is owed (the liability). A bank's balance sheet operates in much the same way. A bank's net worth is also referred to as **bank capital**. A bank has assets such as cash held in its vaults, monies that the bank holds at the Federal Reserve bank (called "reserves"), loans that are made to customers, and bonds.

[\[link\]](#) illustrates a hypothetical and simplified balance sheet for the Safe and Secure Bank. Because of the two-column format of the balance sheet, with the T-shape formed by the vertical line down the middle and the horizontal line under "Assets" and "Liabilities," it is sometimes called a **T-account**.

A Balance Sheet for the Safe and Secure Bank

Assets		Liabilities + Net Worth	
Loans	\$5 million	Deposits	\$10 million
U.S. Government Securities (USGS)	\$4 million		
Reserves	\$2 million	Net Worth	\$1 million

The "T" in a T-account separates the assets of a firm, on the left, from its liabilities, on the right. All firms use T-accounts, though most are much more complex. For a bank, the assets are the financial instruments that either the bank is holding (its reserves) or those instruments where other parties owe money to the bank—like loans made by the bank and U.S. Government Securities, such as U.S. treasury bonds purchased by the bank. Liabilities are what the bank owes to others. Specifically, the bank owes any deposits made in the bank to those who have made them. The net worth of the bank is the total assets minus total liabilities. Net worth is included

on the liabilities side to have the T account balance to zero. For a healthy business, net worth will be positive. For a bankrupt firm, net worth will be negative. In either case, on a bank's T-account, assets will always equal liabilities plus net worth.

When bank customers deposit money into a checking account, savings account, or a certificate of deposit, the bank views these deposits as liabilities. After all, the bank owes these deposits to its customers, when the customers wish to withdraw their money. In the example shown in [\[link\]](#), the Safe and Secure Bank holds \$10 million in deposits.

Loans are the first category of bank assets shown in [\[link\]](#). Say that a family takes out a 30-year mortgage loan to purchase a house, which means that the borrower will repay the loan over the next 30 years. This loan is clearly an asset from the bank's perspective, because the borrower has a legal obligation to make payments to the bank over time. But in practical terms, how can the value of the mortgage loan that is being paid over 30 years be measured in the present? One way of measuring the value of something—whether a loan or anything else—is by estimating what another party in the market is willing to pay for it. Many banks issue home loans, and charge various handling and processing fees for doing so, but then sell the loans to other banks or financial institutions who collect the loan payments. The market where loans are made to borrowers is called the primary loan market, while the market in which these loans are bought and sold by financial institutions is the secondary loan market.

One key factor that affects what financial institutions are willing to pay for a loan, when they buy it in the secondary loan market, is the perceived riskiness of the loan: that is, given the characteristics of the borrower, such as income level and whether the local economy is performing strongly, what proportion of loans of this type will be repaid? The greater the risk that a loan will not be repaid, the less that any financial institution will pay to acquire the loan. Another key factor is to compare the interest rate charged on the original loan with the current interest rate in the economy. If the original loan made at some point in the past requires the borrower to pay a low interest rate, but current interest rates are relatively high, then a financial institution will pay less to acquire the loan. In contrast, if the

original loan requires the borrower to pay a high interest rate, while current interest rates are relatively low, then a financial institution will pay more to acquire the loan. For the Safe and Secure Bank in this example, the total value of its loans if they were sold to other financial institutions in the secondary market is \$5 million.

The second category of bank asset is bonds, which are a common mechanism for borrowing, used by the federal and local government, and also private companies, and nonprofit organizations. A bank takes some of the money it has received in deposits and uses the money to buy bonds—typically bonds issued by the U.S. government. Government bonds are low-risk because the government is virtually certain to pay off the bond, albeit at a low rate of interest. These bonds are an asset for banks in the same way that loans are an asset: The bank will receive a stream of payments in the future. In our example, the Safe and Secure Bank holds bonds worth a total value of \$4 million.

The final entry under assets is **reserves**, which is money that the bank keeps on hand, and that is not loaned out or invested in bonds—and thus does not lead to interest payments. The Federal Reserve requires that banks keep a certain percentage of depositors' money on "reserve," which means either in their vaults or kept at the Federal Reserve Bank. This is called a reserve requirement. ([Monetary Policy and Bank Regulation](#) will explain how the level of these required reserves are one policy tool that governments have to influence bank behavior.) Additionally, banks may also want to keep a certain amount of reserves on hand in excess of what is required. The Safe and Secure Bank is holding \$2 million in reserves.

The net worth of a bank is defined as its total assets minus its total liabilities. For the Safe and Secure Bank shown in [\[link\]](#), net worth is equal to \$1 million; that is, \$11 million in assets minus \$10 million in liabilities. For a financially healthy bank, the net worth will be positive. If a bank has negative net worth and depositors tried to withdraw their money, the bank would not be able to give all depositors their money.

How Banks Go Bankrupt

A bank that is bankrupt will have a negative net worth, meaning its assets will be worth less than its liabilities. How can this happen? Again, looking at the balance sheet helps to explain.

A well-run bank will assume that a small percentage of borrowers will not repay their loans on time, or at all, and factor these missing payments into its planning. Remember, the calculations of the expenses of banks every year includes a factor for loans that are not repaid, and the value of a bank's loans on its balance sheet assumes a certain level of riskiness because some loans will not be repaid. Even if a bank expects a certain number of loan defaults, it will suffer if the number of loan defaults is much greater than expected, as can happen during a recession. For example, if the Safe and Secure Bank in [\[link\]](#) experienced a wave of unexpected defaults, so that its loans declined in value from \$5 million to \$3 million, then the assets of the Safe and Secure Bank would decline so that the bank had negative net worth.

Note:

What led to the financial crisis of 2008–2009?

Many banks make mortgage loans so that people can buy a home, but then do not keep the loans on their books as an asset. Instead, the bank sells the loan. These loans are “securitized,” which means that they are bundled together into a financial security that is sold to investors. Investors in these mortgage-backed securities receive a rate of return based on the level of payments that people make on all the mortgages that stand behind the security.

Securitization offers certain advantages. If a bank makes most of its loans in a local area, then the bank may be financially vulnerable if the local economy declines, so that many people are unable to make their payments. But if a bank sells its local loans, and then buys a mortgage-backed security based on home loans in many parts of the country, it can avoid being exposed to local financial risks. (In the simple example in the text, banks just own “bonds.” In reality, banks can own a number of financial instruments, as long as these financial investments are safe enough to satisfy the government bank regulators.) From the standpoint of a local

homebuyer, securitization offers the benefit that a local bank does not need to have lots of extra funds to make a loan, because the bank is only planning to hold that loan for a short time, before selling the loan so that it can be pooled into a financial security.

But securitization also offers one potentially large disadvantage. If a bank is going to hold a mortgage loan as an asset, the bank has an incentive to scrutinize the borrower carefully to ensure that the loan is likely to be repaid. However, a bank that is going to sell the loan may be less careful in making the loan in the first place. The bank will be more willing to make what are called “subprime loans,” which are loans that have characteristics like low or zero down-payment, little scrutiny of whether the borrower has a reliable income, and sometimes low payments for the first year or two that will be followed by much higher payments after that. Some subprime loans made in the mid-2000s were later dubbed NINJA loans: loans made even though the borrower had demonstrated No Income, No Job, or Assets. These subprime loans were typically sold and turned into financial securities—but with a twist. The idea was that if losses occurred on these mortgage-backed securities, certain investors would agree to take the first, say, 5% of such losses. Other investors would agree to take, say, the next 5% of losses. By this approach, still other investors would not need to take any losses unless these mortgage-backed financial securities lost 25% or 30% or more of their total value. These complex securities, along with other economic factors, encouraged a large expansion of subprime loans in the mid-2000s.

The economic stage was now set for a banking crisis. Banks thought they were buying only ultra-safe securities, because even though the securities were ultimately backed by risky subprime mortgages, the banks only invested in the part of those securities where they were protected from small or moderate levels of losses. But as housing prices fell after 2007, and the deepening recession made it harder for many people to make their mortgage payments, many banks found that their mortgage-backed financial assets could end up being worth much less than they had expected—and so the banks were staring bankruptcy in the face. In the 2008–2011 period, 318 banks failed in the United States.

The risk of an unexpectedly high level of loan defaults can be especially difficult for banks because a bank's liabilities, namely the deposits of its customers, can be withdrawn quickly, but many of the bank's assets like loans and bonds will only be repaid over years or even decades. This **asset-liability time mismatch**—a bank's liabilities can be withdrawn in the short term while its assets are repaid in the long term—can cause severe problems for a bank. For example, imagine a bank that has loaned a substantial amount of money at a certain interest rate, but then sees interest rates rise substantially. The bank can find itself in a precarious situation. If it does not raise the interest rate it pays to depositors, then deposits will flow to other institutions that offer the higher interest rates that are now prevailing. However, if the bank raises the interest rates that it pays to depositors, it may end up in a situation where it is paying a higher interest rate to depositors than it is collecting from those past loans that were made at lower interest rates. Clearly, the bank cannot survive in the long term if it is paying out more in interest to depositors than it is receiving from borrowers.

How can banks protect themselves against an unexpectedly high rate of loan defaults and against the risk of an asset-liability time mismatch? One strategy is for a bank to **diversify** its loans, which means lending to a variety of customers. For example, suppose a bank specialized in lending to a niche market—say, making a high proportion of its loans to construction companies that build offices in one downtown area. If that one area suffers an unexpected economic downturn, the bank will suffer large losses. However, if a bank loans both to consumers who are buying homes and cars and also to a wide range of firms in many industries and geographic areas, the bank is less exposed to risk. When a bank diversifies its loans, those categories of borrowers who have an unexpectedly large number of defaults will tend to be balanced out, according to random chance, by other borrowers who have an unexpectedly low number of defaults. Thus, diversification of loans can help banks to keep a positive net worth. However, if a widespread recession occurs that touches many industries and geographic areas, diversification will not help.

Along with diversifying their loans, banks have several other strategies to reduce the risk of an unexpectedly large number of loan defaults. For

example, banks can sell some of the loans they make in the secondary loan market, as described earlier, and instead hold a greater share of assets in the form of government bonds or reserves. Nevertheless, in a lengthy recession, most banks will see their net worth decline because a higher share of loans will not be repaid in tough economic times.

Key Concepts and Summary

Banks facilitate the use of money for transactions in the economy because people and firms can use bank accounts when selling or buying goods and services, when paying a worker or being paid, and when saving money or receiving a loan. In the financial capital market, banks are financial intermediaries; that is, they operate between savers who supply financial capital and borrowers who demand loans. A balance sheet (sometimes called a T-account) is an accounting tool which lists assets in one column and liabilities in another column. The liabilities of a bank are its deposits. The assets of a bank include its loans, its ownership of bonds, and its reserves (which are not loaned out). The net worth of a bank is calculated by subtracting the bank's liabilities from its assets. Banks run a risk of negative net worth if the value of their assets declines. The value of assets can decline because of an unexpectedly high number of defaults on loans, or if interest rates rise and the bank suffers an asset-liability time mismatch in which the bank is receiving a low rate of interest on its long-term loans but must pay the currently higher market rate of interest to attract depositors. Banks can protect themselves against these risks by choosing to diversify their loans or to hold a greater proportion of their assets in bonds and reserves. If banks hold only a fraction of their deposits as reserves, then the process of banks' lending money, those loans being re-deposited in banks, and the banks making additional loans will create money in the economy.

Self-Check Questions

Exercise:

Problem:

Explain why the money listed under assets on a bank balance sheet may not actually be in the bank?

Solution:

A bank's assets include cash held in their vaults, but assets also include monies that the bank holds at the Federal Reserve Bank (called "reserves"), loans that are made to customers, and bonds.

Review Questions**Exercise:**

Problem: Why is a bank called a financial intermediary?

Exercise:

Problem: What does a balance sheet show?

Exercise:

Problem: What are the assets of a bank? What are its liabilities?

Exercise:

Problem: How do you calculate the net worth of a bank?

Exercise:

Problem: How can a bank end up with negative net worth?

Exercise:

Problem: What is the asset-liability time mismatch that all banks face?

Exercise:

Problem: What is the risk if a bank does not diversify its loans?

Critical Thinking Questions

Exercise:

Problem:

Explain the difference between how you would characterize bank deposits and loans as assets and liabilities on your own personal balance sheet and how a bank would characterize deposits and loans as assets and liabilities on its balance sheet.

Problems

Exercise:

Problem:

A bank has deposits of \$400. It holds reserves of \$50. It has purchased government bonds worth \$70. It has made loans of \$500. Set up a T-account balance sheet for the bank, with assets and liabilities, and calculate the bank's net worth.

References

Credit Union National Association. 2014. "Monthly Credit Union Estimates." Last accessed March 4, 2015. <http://www.cuna.org/Research-And-Strategy/Credit-Union-Data-And-Statistics/>.

Dallas Federal Reserve. 2013. "Ending 'Too Big To Fail': A Proposal for Reform Before It's Too Late". Accessed March 4, 2015. <http://www.dallasfed.org/news/speeches/fisher/2013/fs130116.cfm>.

Richard W. Fisher. “Ending 'Too Big to Fail': A Proposal for Reform Before It's Too Late (With Reference to Patrick Henry, Complexity and Reality) Remarks before the Committee for the Republic, Washington, D.C. Dallas Federal Reserve. January 16, 2013.

“Commercial Banks in the U.S.” Federal Reserve Bank of St. Louis. Accessed November 2013.
<http://research.stlouisfed.org/fred2/series/USNUM>.

Glossary

asset

item of value owned by a firm or an individual

asset–liability time mismatch

a bank’s liabilities can be withdrawn in the short term while its assets are repaid in the long term

balance sheet

an accounting tool that lists assets and liabilities

bank capital

a bank’s net worth

depository institution

institution that accepts money deposits and then uses these to make loans

diversify

making loans or investments with a variety of firms, to reduce the risk of being adversely affected by events at one or a few firms

financial intermediary

an institution that operates between a saver with financial assets to invest and an entity who will borrow those assets and pay a rate of return

liability

any amount or debt owed by a firm or an individual

net worth

the excess of the asset value over and above the amount of the liability;
total assets minus total liabilities

payment system

helps an economy exchange goods and services for money or other
financial assets

reserves

funds that a bank keeps on hand and that are not loaned out or invested
in bonds

T-account

a balance sheet with a two-column format, with the T-shape formed by
the vertical line down the middle and the horizontal line under the
column headings for “Assets” and “Liabilities”

transaction costs

the costs associated with finding a lender or a borrower for money

How Banks Create Money

By the end of this section, you will be able to:

- Utilize the money multiplier formula to determine how banks create money
- Analyze and create T-account balance sheets
- Evaluate the risks and benefits of money and banks

Banks and money are intertwined. It is not just that most money is in the form of bank accounts. The banking system can literally create money through the process of making loans. Let's see how.

Money Creation by a Single Bank

Start with a hypothetical bank called Singleton Bank. The bank has \$10 million in deposits. The T-account balance sheet for Singleton Bank, when it holds all of the deposits in its vaults, is shown in [\[link\]](#). At this stage, Singleton Bank is simply storing money for depositors and is using these deposits to make loans. In this simplified example, Singleton Bank cannot earn any interest income from these loans and cannot pay its depositors an interest rate either.

Singleton Bank's Balance Sheet: Receives \$10 million in Deposits

Assets		Liabilities + Net Worth	
Reserves	\$10 million	Deposits	\$10 million

Singleton Bank is required by the Federal Reserve to keep \$1 million on reserve (10% of total deposits). It will loan out the remaining \$9 million. By loaning out the \$9 million and charging interest, it will be able to make interest payments to depositors and earn interest income for Singleton Bank (for now, we will keep it simple and not put interest income on the balance sheet). Instead of becoming just a storage place for deposits, Singleton Bank can become a financial intermediary between savers and borrowers.

This change in business plan alters Singleton Bank's balance sheet, as shown in [\[link\]](#). Singleton's assets have changed; it now has \$1 million in reserves and a loan to Hank's Auto Supply of \$9 million. The bank still has \$10 million in deposits.

Singleton Bank's Balance Sheet: 10% Reserves, One Round of Loans

Assets		Liabilities + Net Worth	
Reserves	\$1 million	Deposits	\$10 million
Loan to Hank's Auto Supply	\$9 million		

Singleton Bank lends \$9 million to Hank's Auto Supply. The bank records this loan by making an entry on the balance sheet to indicate that a loan has been made. This loan is an asset, because it will generate interest income for the bank. Of course, the loan officer is not going to let Hank walk out of the bank with \$9 million in cash. The bank issues Hank's Auto Supply a cashier's check for the \$9 million. Hank deposits the loan in his regular

checking account with First National. The deposits at First National rise by \$9 million and its reserves also rise by \$9 million, as [\[link\]](#) shows. First National must hold 10% of additional deposits as required reserves but is free to loan out the rest

First National Balance Sheet

Assets		Liabilities + Net Worth	
Reserves	+ \$9 million	Deposits	+ \$9 million

Making loans that are deposited into a demand deposit account increases the M1 money supply. Remember the definition of M1 includes checkable (demand) deposits, which can be easily used as a medium of exchange to buy goods and services. Notice that the money supply is now \$19 million: \$10 million in deposits in Singleton bank and \$9 million in deposits at First National. Obviously these deposits will be drawn down as Hank's Auto Supply writes checks to pay its bills. But the bigger picture is that a bank must hold enough money in reserves to meet its liabilities; the rest the bank loans out. In this example so far, bank lending has expanded the money supply by \$9 million.

Now, First National must hold only 10% as required reserves (\$900,000) but can lend out the other 90% (\$8.1 million) in a loan to Jack's Chevy Dealership as shown in [\[link\]](#).

First National Balance Sheet

Assets		Liabilities + Net Worth	
Reserves	\$900,000	Deposits	+ \$9 million
Loans	\$8.1 million		

If Jack's deposits the loan in its checking account at Second National, the money supply just increased by an additional \$8.1 million, as [\[link\]](#) shows.

Second National Bank's Balance Sheet

Assets		Liabilities + Net Worth	
Reserves	+ \$8.1 million	Deposits	+ \$8.1 million

How is this money creation possible? It is possible because there are multiple banks in the financial system, they are required to hold only a fraction of their deposits, and loans end up deposited in other banks, which increases deposits and, in essence, the money supply.

Note: Watch this [video](#) to learn more about how banks create money.

The Money Multiplier and a Multi-Bank System

In a system with multiple banks, the initial excess reserve amount that Singleton Bank decided to lend to Hank's Auto Supply was deposited into First National Bank, which is free to loan out \$8.1 million. If all banks loan out their excess reserves, the money supply

will expand. In a multi-bank system, the amount of money that the system can create is found by using the money multiplier. The money multiplier tells us by how many times a loan will be “multiplied” as it is spent in the economy and then re-deposited in other banks.

Fortunately, a formula exists for calculating the total of these many rounds of lending in a banking system. The **money multiplier formula** is:

Equation:

$$\frac{1}{\text{Reserve Requirement}}$$

The money multiplier is then multiplied by the change in excess reserves to determine the total amount of M1 money supply created in the banking system. See the Work it Out feature to walk through the multiplier calculation.

Note:

Using the Money Multiplier Formula

Using the money multiplier for the example in this text:

Step 1. In the case of Singleton Bank, for whom the reserve requirement is 10% (or 0.10), the money multiplier is 1 divided by .10, which is equal to 10.

Step 2. We have identified that the excess reserves are \$9 million, so, using the formula we can determine the total change in the M1 money supply:

Equation:

$$\begin{aligned}\text{Total Change in the M1 Money Supply} &= \frac{1}{\text{Reserve Requirement}} \times \text{Excess Requirement} \\ &= \frac{1}{0.10} \times \$9 \text{ million} \\ &= 10 \times \$9 \text{ million} \\ &= \$90 \text{ million}\end{aligned}$$

Step 3. Thus, we can say that, in this example, the total quantity of money generated in this economy after all rounds of lending are completed will be \$90 million.

Cautions about the Money Multiplier

The money multiplier will depend on the proportion of reserves that banks are required to hold by the Federal Reserve Bank. Additionally, a bank can also choose to hold extra reserves. Banks may decide to vary how much they hold in reserves for two reasons: macroeconomic conditions and government rules. When an economy is in recession, banks are likely to hold a higher proportion of reserves because they fear that loans are less likely

to be repaid when the economy is slow. The Federal Reserve may also raise or lower the required reserves held by banks as a policy move to affect the quantity of money in an economy, as [Monetary Policy and Bank Regulation](#) will discuss.

The process of how banks create money shows how the quantity of money in an economy is closely linked to the quantity of lending or credit in the economy. Indeed, all of the money in the economy, except for the original reserves, is a result of bank loans that are re-deposited and loaned out, again, and again.

Finally, the money multiplier depends on people re-depositing the money that they receive in the banking system. If people instead store their cash in safe-deposit boxes or in shoeboxes hidden in their closets, then banks cannot recirculate the money in the form of loans. Indeed, central banks have an incentive to assure that bank deposits are safe because if people worry that they may lose their bank deposits, they may start holding more money in cash, instead of depositing it in banks, and the quantity of loans in an economy will decline. Low-income countries have what economists sometimes refer to as “mattress savings,” or money that people are hiding in their homes because they do not trust banks. When mattress savings in an economy are substantial, banks cannot lend out those funds and the money multiplier cannot operate as effectively. The overall quantity of money and loans in such an economy will decline.

Money and Banks—Benefits and Dangers

Money and banks are marvelous social inventions that help a modern economy to function. Compared with the alternative of barter, money makes market exchanges vastly easier in goods, labor, and financial markets. Banking makes money still more effective in facilitating exchanges in goods and labor markets. Moreover, the process of banks making loans in financial capital markets is intimately tied to the creation of money.

But the extraordinary economic gains that are possible through money and banking also suggest some possible corresponding dangers. If banks are not working well, it sets off a decline in convenience and safety of transactions throughout the economy. If the banks are under financial stress, because of a widespread decline in the value of their assets, loans may become far less available, which can deal a crushing blow to sectors of the economy that depend on borrowed money like business investment, home construction, and car manufacturing. The Great Recession of 2008–2009 illustrated this pattern.

Note:

The Many Disguises of Money: From Cowries to Bit Coins

The global economy has come a long way since it started using cowrie shells as currency. We have moved away from commodity and commodity-backed paper money to fiat currency. As technology and global integration increases, the need for paper currency is diminishing, too. Every day, we witness the increased use of debit and credit cards.

The latest creation and perhaps one of the purest forms of fiat money is the Bitcoin. Bitcoins are a digital currency that allows users to buy goods and services online. Products and services such as videos and books may be purchased using Bitcoins. It is not backed by any commodity nor has it been decreed by any government as legal tender, yet it is used as a medium of exchange and its value (online at least) can be stored. It is also unregulated by any central bank, but is created online through people solving very complicated mathematics problems and getting paid afterward. Bitcoin.org is an information source if you are curious. Bitcoins are a relatively new type of money. At present, because it is not sanctioned as a legal currency by any country nor regulated by any central bank, it lends itself for use in illegal trading activities as well as legal ones. As technology increases and the need to reduce transactions costs associated with using traditional forms of money increases, Bitcoins or some sort of digital currency may replace our dollar bill, just as the cowrie shell was replaced.

Key Concepts and Summary

The money multiplier is defined as the quantity of money that the banking system can generate from each \$1 of bank reserves. The formula for calculating the multiplier is $1/\text{reserve ratio}$, where the reserve ratio is the fraction of deposits that the bank wishes to hold as reserves. The quantity of money in an economy and the quantity of credit for loans are inextricably intertwined. Much of the money in an economy is created by the network of banks making loans, people making deposits, and banks making more loans.

Given the macroeconomic dangers of a malfunctioning banking system, [Monetary Policy and Bank Regulation](#) will discuss government policies for controlling the money supply and for keeping the banking system safe.

Self-Check Questions

Exercise:

Problem:

Imagine that you are in the position of buying loans in the secondary market (that is, buying the right to collect the payments on loans made by banks) for a bank or other financial services company. Explain why you would be willing to pay more or less for a given loan if:

- The borrower has been late on a number of loan payments
 - Interest rates in the economy as a whole have risen since the loan was made
 - The borrower is a firm that has just declared a high level of profits
 - Interest rates in the economy as a whole have fallen since the loan was made
-

Solution:

- a. A borrower who has been late on a number of loan payments looks perhaps less likely to repay the loan, or to repay it on time, and so you would want to pay less for that loan.
- b. If interest rates generally have risen, then this loan made at a time of relatively lower interest rates looks less attractive, and you would pay less for it.
- c. If the borrower is a firm with a record of high profits, then it is likely to be able to repay the loan, and you would be willing to pay more for the loan.
- d. If interest rates in the economy have fallen, then the loan is worth more.

Review Questions**Exercise:**

Problem: How do banks create money?

Exercise:

Problem: What is the formula for the money multiplier?

Critical Thinking Questions**Exercise:**

Problem: Should banks have to hold 100% of their deposits? Why or why not?

Exercise:**Problem:**

Explain what will happen to the money multiplier process if there is an increase in the reserve requirement?

Exercise:**Problem:**

What do you think the Federal Reserve Bank did to the reserve requirement during the Great Recession of 2008–2009?

Problems

Exercise:

Problem:

Humongous Bank is the only bank in the economy. The people in this economy have \$20 million in money, and they deposit all their money in Humongous Bank.

- a. Humongous Bank decides on a policy of holding 100% reserves. Draw a T-account for the bank.
- b. Humongous Bank is required to hold 5% of its existing \$20 million as reserves, and to loan out the rest. Draw a T-account for the bank after this first round of loans has been made.
- c. Assume that Humongous bank is part of a multibank system. How much will money supply increase with that original loan of \$19 million?

References

Bitcoin. 2013. www.bitcoin.org.

National Public Radio. *Lawmakers and Regulators Take Closer Look at Bitcoin*. November 19, 2013. <http://thedianerehmshow.org/shows/2013-11-19/lawmakers-and-regulators-take-closer-look-bitcoin>.

Glossary

money multiplier formula

total money in the economy divided by the original quantity of money, or change in the total money in the economy divided by a change in the original quantity of money

Introduction to Monetary Policy and Bank Regulation

class="introduction"

Marriner S. Eccles Federal Reserve Headquarters, Washington D.C.

Some of the most
influential
decisions
regarding
monetary policy in
the United States
are made behind
these doors.

(Credit:
modification of
work by
“squirrel83”/Flick
r Creative
Commons)



Note:**The Problem of the Zero Percent Interest Rate Lower Bound**

Most economists believe that monetary policy (the manipulation of interest rates and credit conditions by a nation's central bank) has a powerful influence on a nation's economy. Monetary policy works when the central bank reduces interest rates and makes credit more available. As a result, business investment and other types of spending increase, causing GDP and employment to grow.

But what if the interest rates banks pay are close to zero already? They cannot be made negative, can they? That would mean that lenders pay borrowers for the privilege of taking their money. Yet, this was the situation the U.S. Federal Reserve found itself in at the end of the 2008–2009 recession. The federal funds rate, which is the interest rate for banks that the Federal Reserve targets with its monetary policy, was slightly above 5% in 2007. By 2009, it had fallen to 0.16%.

The Federal Reserve's situation was further complicated because fiscal policy, the other major tool for managing the economy, was constrained by fears that the federal budget deficit and the public debt were already too high. What were the Federal Reserve's options? How could monetary policy be used to stimulate the economy? The answer, as we will see in this chapter, was to change the rules of the game.

Note:**Introduction to Monetary Policy and Bank Regulation**

In this chapter, you will learn about:

- The Federal Reserve Banking System and Central Banks
- Bank Regulation
- How a Central Bank Executes Monetary Policy
- Monetary Policy and Economic Outcomes
- Pitfalls for Monetary Policy

Money, loans, and banks are all tied together. Money is deposited in bank accounts, which is then loaned to businesses, individuals, and other banks. When the interlocking system of money, loans, and banks works well, economic transactions are made smoothly in goods and labor markets and savers are connected with borrowers. If the money and banking system does not operate smoothly, the economy can either fall into recession or suffer prolonged inflation.

The government of every country has public policies that support the system of money, loans, and banking. But these policies do not always work perfectly. This chapter discusses how monetary policy works and what may prevent it from working perfectly.

The Federal Reserve Banking System and Central Banks

By the end of this section, you will be able to:

- Explain the structure and organization of the U.S. Federal Reserve
- Discuss how central banks impact monetary policy, promote financial stability, and provide banking services

In making decisions about the money supply, a central bank decides whether to raise or lower interest rates and, in this way, to influence macroeconomic policy, whose goal is low unemployment and low inflation. The central bank is also responsible for regulating all or part of the nation's banking system to protect bank depositors and insure the health of the bank's balance sheet.

The organization responsible for conducting monetary policy and ensuring that a nation's financial system operates smoothly is called the **central bank**. Most nations have central banks or currency boards. Some prominent central banks around the world include the European Central Bank, the Bank of Japan, and the Bank of England. In the United States, the central bank is called the Federal Reserve—often abbreviated as just “the Fed.” This section explains the organization of the U.S. Federal Reserve and identifies the major responsibilities of a central bank.

Structure/Organization of the Federal Reserve

Unlike most central banks, the Federal Reserve is semi-decentralized, mixing government appointees with representation from private-sector banks. At the national level, it is run by a Board of Governors, consisting of seven members appointed by the President of the United States and confirmed by the Senate. Appointments are for 14-year terms and they are arranged so that one term expires January 31 of every even-numbered year. The purpose of the long and staggered terms is to insulate the Board of Governors as much as possible from political pressure so that policy decisions can be made based only on their economic merits. Additionally, except when filling an unfinished term, each member only serves one term, further insulating decision-making from politics. Policy decisions of the

Fed do not require congressional approval, and the President cannot ask for the resignation of a Federal Reserve Governor as the President can with cabinet positions.

One member of the Board of Governors is designated as the Chair. For example, from 1987 until early 2006, the Chair was Alan Greenspan. From 2006 until 2014, Ben Bernanke held the post. The current Chair, Janet Yellen, has made many headlines already. Why? See the following Clear It Up feature to find out.

Note:

Who has the most immediate economic power in the world?
Chair of the Federal Reserve Board



Janet L. Yellen is the first woman to hold the position of Chair of the Federal Reserve Board of Governors.
(Credit: Board of Governors of the Federal Reserve System)

What individual can make financial market crash or soar just by making a public statement? It is not Bill Gates or Warren Buffett. It is not even the President of the United States. The answer is the Chair of the Federal Reserve Board of Governors. In early 2014, Janet L. Yellen, shown in [\[link\]](#) became the first woman to hold this post. Yellen has been described in the media as “perhaps the most qualified Fed chair in history.” With a Ph.D. in economics from Yale University, Yellen has taught macroeconomics at Harvard, the London School of Economics, and most recently at the University of California at Berkeley. From 2004–2010, Yellen was President of the Federal Reserve Bank of San Francisco. Not an ivory tower economist, Yellen became one of the few economists who warned about a possible bubble in the housing market, more than two years before the financial crisis occurred. Yellen served on the Board of Governors of the Federal Reserve twice, most recently as Vice Chair. She also spent two years as Chair of the President’s Council of Economic Advisors. If experience and credentials mean anything, Yellen is likely to be an effective Fed chair.

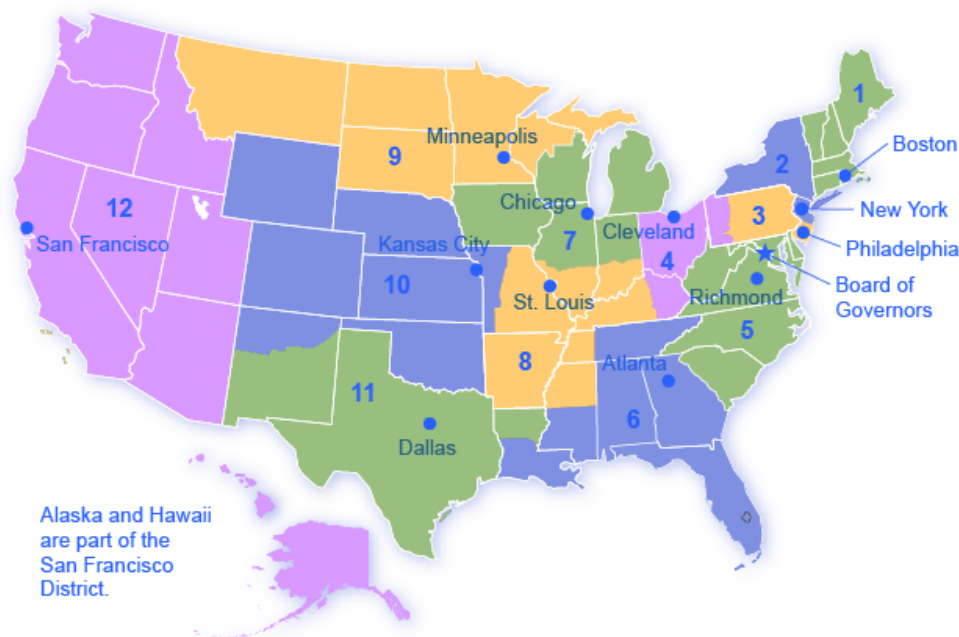
The Fed Chair is first among equals on the Board of Governors. While he or she has only one vote, the Chair controls the agenda, and is the public voice of the Fed, so he or she has more power and influence than one might expect.

Note: Visit this [website](#) to see who the current members of the Federal Reserve Board of Governors are. You can follow the links provided for each board member to learn more about their backgrounds, experiences, and when their terms on the board will end.

The Federal Reserve is more than the Board of Governors. The Fed also includes 12 regional Federal Reserve banks, each of which is responsible

for supporting the commercial banks and economy generally in its district. The Federal Reserve districts and the cities where their regional headquarters are located are shown in [\[link\]](#). The commercial banks in each district elect a Board of Directors for each regional Federal Reserve bank, and that board chooses a president for each regional Federal Reserve district. Thus, the Federal Reserve System includes both federally and private-sector appointed leaders.

The Twelve Federal Reserve Districts



There are twelve regional Federal Reserve banks, each with its district.

What Does a Central Bank Do?

The Federal Reserve, like most central banks, is designed to perform three important functions:

1. To conduct monetary policy
2. To promote stability of the financial system
3. To provide banking services to commercial banks and other depository institutions, and to provide banking services to the federal government.

The first two functions are sufficiently important that we will discuss them in their own modules; the third function we will discuss here.

The Federal Reserve provides many of the same services to banks as banks provide to their customers. For example, all commercial banks have an account at the Fed where they deposit reserves. Similarly, banks can obtain loans from the Fed through the “discount window” facility, which will be discussed in more detail later. The Fed is also responsible for check processing. When you write a check, for example, to buy groceries, the grocery store deposits the check in its bank account. Then, the physical check (or an image of that actual check) is returned to your bank, after which funds are transferred from your bank account to the account of the grocery store. The Fed is responsible for each of these actions.

On a more mundane level, the Federal Reserve ensures that enough currency and coins are circulating through the financial system to meet public demands. For example, each year the Fed increases the amount of currency available in banks around the Christmas shopping season and reduces it again in January.

Finally, the Fed is responsible for assuring that banks are in compliance with a wide variety of consumer protection laws. For example, banks are forbidden from discriminating on the basis of age, race, sex, or marital status. Banks are also required to disclose publicly information about the loans they make for buying houses and how those loans are distributed geographically, as well as by sex and race of the loan applicants.

Key Concepts and Summary

The most prominent task of a central bank is to conduct monetary policy, which involves changes to interest rates and credit conditions, affecting the amount of borrowing and spending in an economy. Some prominent central banks around the world include the U.S. Federal Reserve, the European Central Bank, the Bank of Japan, and the Bank of England.

Self-Check Question

Exercise:**Problem:**

Why is it important for the members of the Board of Governors of the Federal Reserve to have longer terms in office than elected officials, like the President?

Solution:

Longer terms insulate the Board from political forces. Since the presidency can potentially change every four years, the Federal Reserve's independence prevents drastic swings in monetary policy with every new administration and allows policy decisions to be made only on economic grounds.

Review Questions**Exercise:****Problem:**

How is a central bank different from a typical commercial bank?

Exercise:**Problem:**

List the three traditional tools that a central bank has for controlling the money supply.

Critical Thinking Questions**Exercise:**

Problem:

Why do presidents typically reappoint Chairs of the Federal Reserve Board even when they were originally appointed by a president of a different political party?

Exercise:**Problem:**

In what ways might monetary policy be superior to fiscal policy? In what ways might it be inferior?

References

Matthews, Dylan. "Nine amazing facts about Janet Yellen, our next Fed chair." Wonkblog. *The Washington Post*. Posted October 09, 2013.
<http://www.washingtonpost.com/blogs/wonkblog/wp/2013/10/09/nine-amazing-facts-about-janet-yellen-our-next-fed-chair/>.

Appelbaum, Binyamin. "Divining the Regulatory Goals of Fed Rivals." *The New York Times*. Posted August 14, 2013.
<http://www.nytimes.com/2013/08/14/business/economy/careers-of-2-fed-contenders-reveal-little-on-regulatory-approach.html?pagewanted=3>.

Glossary

central bank

institution which conducts a nation's monetary policy and regulates its banking system

Bank Regulation

By the end of this section, you will be able to:

- Discuss the relationship between bank regulation and monetary policy
- Explain bank supervision
- Explain how deposit insurance and lender of last resort are two strategies to protect against bank runs

A safe and stable national financial system is a critical concern of the Federal Reserve. The goal is not only to protect individuals' savings, but to protect the integrity of the financial system itself. This esoteric task is usually behind the scenes, but came into view during the 2008–2009 financial crisis, when for a brief period of time, critical parts of the financial system failed and firms became unable to obtain financing for ordinary parts of their business. Imagine if suddenly you were unable to access the money in your bank accounts because your checks were not accepted for payment and your debit cards were declined. This gives an idea of what a failure of the payments/financial system is like.

Bank regulation is intended to maintain the solvency of banks by avoiding excessive risk. Regulation falls into a number of categories, including reserve requirements, capital requirements, and restrictions on the types of investments banks may make. In [Money and Banking](#), we learned that banks are required to hold a minimum percentage of their deposits on hand as reserves. “On hand” is a bit of a misnomer because, while a portion of bank reserves are held as cash in the bank, the majority are held in the bank's account at the Federal Reserve, and their purpose is to cover desired withdrawals by depositors. Another part of bank regulation is restrictions on the types of investments banks are allowed to make. Banks are allowed to make loans to businesses, individuals, and other banks. They are allowed to purchase U.S. Treasury securities but, to protect depositors, they are not permitted to invest in the stock market or other assets that are perceived as too risky.

Bank capital is the difference between a bank's assets and its liabilities. In other words, it is a bank's net worth. A bank must have positive net worth; otherwise it is insolvent or bankrupt, meaning it would not have enough

assets to pay back its liabilities. Regulation requires that banks maintain a minimum net worth, usually expressed as a percent of their assets, to protect their depositors and other creditors.

Bank Supervision

Several government agencies monitor the balance sheets of banks to make sure they have positive net worth and are not taking too high a level of risk. Within the U.S. Department of the Treasury, the Office of the Comptroller of the Currency has a national staff of bank examiners who conduct on-site reviews of the 1,500 or so of the largest national banks. The bank examiners also review any foreign banks that have branches in the United States. The Office of the Comptroller of the Currency also monitors and regulates about 800 savings and loan institutions.

The National Credit Union Administration (NCUA) supervises credit unions, which are nonprofit banks owned and run by their members. There are over 6,000 credit unions in the U.S. economy, though the typical credit union is small compared to most banks.

The Federal Reserve also has some responsibility for supervising financial institutions. For example, conglomerate firms that own banks and other businesses are called “bank holding companies.” While other regulators like the Office of the Comptroller of the Currency supervises the banks, the Federal Reserve supervises the holding companies.

When the supervision of banks (and bank-like institutions such as savings and loans and credit unions) works well, most banks will remain financially healthy most of the time. If the bank supervisors find that a bank has low or negative net worth, or is making too high a proportion of risky loans, they can require that the bank change its behavior—or, in extreme cases, even force the bank to be closed or sold to a financially healthy bank.

Bank supervision can run into both practical and political questions. The practical question is that measuring the value of a bank’s assets is not always straightforward. As discussed in [Money and Banking](#), a bank’s assets are its loans, and the value of these assets depends on estimates about

the risk that these loans will not be repaid. These issues can become even more complex when a bank makes loans to banks or firms in other countries, or arranges financial deals that are much more complex than a basic loan.

The political question arises because the decision by a bank supervisor to require a bank to close or to change its financial investments is often controversial, and the bank supervisor often comes under political pressure from the owners of the bank and the local politicians to keep quiet and back off.

For example, many observers have pointed out that Japan's banks were in deep financial trouble through most of the 1990s; however, nothing substantial had been done about it by the early 2000s. A similar unwillingness to confront problems with struggling banks is visible across the rest of the world, in East Asia, Latin America, Eastern Europe, Russia, and elsewhere.

In the United States, laws were passed in the 1990s requiring that bank supervisors make their findings open and public, and that they act as soon as a problem is identified. However, as many U.S. banks were staggered by the recession of 2008–2009, critics of the bank regulators asked pointed questions about why the regulators had not foreseen the financial shakiness of the banks earlier, before such large losses had a chance to accumulate.

Bank Runs

Back in the nineteenth century and during the first few decades of the twentieth century (around and during the Great Depression), putting your money in a bank could be nerve-wracking. Imagine that the net worth of your bank became negative, so that the bank's assets were not enough to cover its liabilities. In this situation, whoever withdrew their deposits first received all of their money, and those who did not rush to the bank quickly enough, lost their money. Depositors racing to the bank to withdraw their deposits, as shown in [\[link\]](#) is called a **bank run**. In the movie *It's a Wonderful Life*, the bank manager, played by Jimmy Stewart, faces a mob of worried bank depositors who want to withdraw their money, but manages

to allay their fears by allowing some of them to withdraw a portion of their deposits—using the money from his own pocket that was supposed to pay for his honeymoon.

A Run on the Bank



Bank runs during the Great Depression only served to worsen the economic situation. (Credit: National Archives and Records Administration)

The risk of bank runs created instability in the banking system. Even a rumor that a bank might experience negative net worth could trigger a bank run and, in a bank run, even healthy banks could be destroyed. Because a bank loans out most of the money it receives, and because it keeps only limited reserves on hand, a bank run of any size would quickly drain any of the bank's available cash. When the bank had no cash remaining, it only intensified the fears of remaining depositors that they could lose their money. Moreover, a bank run at one bank often triggered a chain reaction of runs on other banks. In the late nineteenth and early twentieth century, bank runs were typically not the original cause of a recession—but they could make a recession much worse.

Deposit Insurance

To protect against bank runs, Congress has put two strategies into place: **deposit insurance** and the lender of last resort. Deposit insurance is an insurance system that makes sure depositors in a bank do not lose their money, even if the bank goes bankrupt. About 70 countries around the world, including all of the major economies, have deposit insurance programs. In the United States, the Federal Deposit Insurance Corporation (FDIC) is responsible for deposit insurance. Banks pay an insurance premium to the FDIC. The insurance premium is based on the bank's level of deposits, and then adjusted according to the riskiness of a bank's financial situation. In 2009, for example, a fairly safe bank with a high net worth might have paid 10–20 cents in insurance premiums for every \$100 in bank deposits, while a risky bank with very low net worth might have paid 50–60 cents for every \$100 in bank deposits.

Bank examiners from the FDIC evaluate the balance sheets of banks, looking at the value of assets and liabilities, to determine the level of riskiness. The FDIC provides deposit insurance for about 6,509 banks (as of the end of 2014). Even if a bank fails, the government guarantees that depositors will receive up to \$250,000 of their money in each account, which is enough for almost all individuals, although not sufficient for many businesses. Since the United States enacted deposit insurance in the 1930s, no one has lost any of their insured deposits. Bank runs no longer happen at insured banks.

Lender of Last Resort

The problem with bank runs is not that insolvent banks will fail; they are, after all, bankrupt and need to be shut down. The problem is that bank runs can cause solvent banks to fail and spread to the rest of the financial system. To prevent this, the Fed stands ready to lend to banks and other financial institutions when they cannot obtain funds from anywhere else. This is known as the **lender of last resort** role. For banks, the central bank acting as a lender of last resort helps to reinforce the effect of deposit insurance and to reassure bank customers that they will not lose their money.

The lender of last resort task can come up in other financial crises, as well. During the panic of the stock market crash in 1987, when the value of U.S. stocks fell by 25% in a single day, the Federal Reserve made a number of short-term emergency loans so that the financial system could keep functioning. During the recession of 2008–2009, the “quantitative easing” policies (discussed below) of the Federal Reserve can be interpreted as a willingness to make short-term credit available as needed in a time when the banking and financial system was under stress.

Key Concepts and Summary

A bank run occurs when there are rumors (possibly true, possibly false) that a bank is at financial risk of having negative net worth. As a result, depositors rush to the bank to withdraw their money and put it someplace safer. Even false rumors, if they cause a bank run, can force a healthy bank to lose its deposits and be forced to close. Deposit insurance guarantees bank depositors that, even if the bank has negative net worth, their deposits will be protected. In the United States, the Federal Deposit Insurance Corporation (FDIC) collects deposit insurance premiums from banks and guarantees bank deposits up to \$250,000. Bank supervision involves inspecting the balance sheets of banks to make sure that they have positive net worth and that their assets are not too risky. In the United States, the Office of the Comptroller of the Currency (OCC) is responsible for supervising banks and inspecting savings and loans and the National Credit Union Administration (NCUA) is responsible for inspecting credit unions. The FDIC and the Federal Reserve also play a role in bank supervision.

When a central bank acts as a lender of last resort, it makes short-term loans available in situations of severe financial panic or stress. The failure of a single bank can be treated like any other business failure. Yet if many banks fail, it can reduce aggregate demand in a way that can bring on or deepen a recession. The combination of deposit insurance, bank supervision, and lender of last resort policies help to prevent weaknesses in the banking system from causing recessions.

Self-Check Questions

Exercise:**Problem:**

Given the danger of bank runs, why do banks not keep the majority of deposits on hand to meet the demands of depositors?

Solution:

Banks make their money from issuing loans and charging interest. The more money that is stored in the bank's vault, the less is available for lending and the less money the bank stands to make.

Exercise:**Problem:**

Bank runs are often described as “self-fulfilling prophecies.” Why is this phrase appropriate to bank runs?

Solution:

The fear and uncertainty created by the suggestion that a bank might fail can lead depositors to withdraw their money. If many depositors do this at the same time, the bank may not be able to meet their demands and will, indeed, fail.

Review Questions**Exercise:****Problem:**

How is bank regulation linked to the conduct of monetary policy?

Exercise:

Problem: What is a bank run?

Exercise:

Problem:

In a program of deposit insurance as it is operated in the United States, what is being insured and who pays the insurance premiums?

Exercise:**Problem:**

In government programs of bank supervision, what is being supervised?

Exercise:

Problem: What is the lender of last resort?

Exercise:**Problem:**

Name and briefly describe the responsibilities of each of the following agencies: FDIC, NCUA, and OCC.

Critical Thinking Question**Exercise:****Problem:**

The term “moral hazard” describes increases in risky behavior resulting from efforts to make that behavior safer. How does the concept of moral hazard apply to deposit insurance and other bank regulations?

References

U.S. Department of the Treasury. “Office of the Comptroller of the Currency.” Accessed November 2013. <http://www.occ.gov/>.

National Credit Union Administration. "About NCUA." Accessed November 2013. <http://www.ncua.gov/about/Pages/default.aspx>.

Glossary

bank run

when depositors race to the bank to withdraw their deposits for fear that otherwise they would be lost

deposit insurance

an insurance system that makes sure depositors in a bank do not lose their money, even if the bank goes bankrupt

lender of last resort

an institution that provides short-term emergency loans in conditions of financial crisis

How a Central Bank Executes Monetary Policy

By the end of this section, you will be able to:

- Explain the reason for open market operations
- Evaluate reserve requirements and discount rates
- Interpret and show bank activity through balance sheets

The most important function of the Federal Reserve is to conduct the nation's monetary policy. Article I, Section 8 of the U.S. Constitution gives Congress the power “to coin money” and “to regulate the value thereof.” As part of the 1913 legislation that created the Federal Reserve, Congress delegated these powers to the Fed. Monetary policy involves managing interest rates and credit conditions, which influences the level of economic activity, as described in more detail below.

A central bank has three traditional tools to implement monetary policy in the economy:

- Open market operations
- Changing reserve requirements
- Changing the discount rate

In discussing how these three tools work, it is useful to think of the central bank as a “bank for banks”—that is, each private-sector bank has its own account at the central bank. We will discuss each of these monetary policy tools in the sections below.

Open Market Operations

The most commonly used tool of monetary policy in the U.S. is **open market operations**. Open market operations take place when the central bank sells or buys U.S. Treasury bonds in order to influence the quantity of bank reserves and the level of interest rates. The specific interest rate targeted in open market operations is the federal funds rate. The name is a bit of a misnomer since the federal funds rate is the interest rate charged by commercial banks making overnight loans to other banks. As such, it is a

very short term interest rate, but one that reflects credit conditions in financial markets very well.

The **Federal Open Market Committee (FOMC)** makes the decisions regarding these open market operations. The FOMC is made up of the seven members of the Federal Reserve's Board of Governors. It also includes five voting members who are drawn, on a rotating basis, from the regional Federal Reserve Banks. The New York district president is a permanent voting member of the FOMC and the other four spots are filled on a rotating, annual basis, from the other 11 districts. The FOMC typically meets every six weeks, but it can meet more frequently if necessary. The FOMC tries to act by consensus; however, the chairman of the Federal Reserve has traditionally played a very powerful role in defining and shaping that consensus. For the Federal Reserve, and for most central banks, open market operations have, over the last few decades, been the most commonly used tool of monetary policy.

Note: Visit this [website](#) for the Federal Reserve to learn more about current monetary policy.

To understand how open market operations affect the money supply, consider the balance sheet of Happy Bank, displayed in [\[link\]](#). [\[link\]](#) (a) shows that Happy Bank starts with \$460 million in assets, divided among reserves, bonds and loans, and \$400 million in liabilities in the form of deposits, with a net worth of \$60 million. When the central bank purchases \$20 million in bonds from Happy Bank, the bond holdings of Happy Bank fall by \$20 million and the bank's reserves rise by \$20 million, as shown in [\[link\]](#) (b). However, Happy Bank only wants to hold \$40 million in reserves (the quantity of reserves that it started with in [\[link\]](#)) (a), so the bank decides to loan out the extra \$20 million in reserves and its loans rise by \$20 million, as shown in [\[link\]](#) (c). The open market operation by the central bank causes Happy Bank to make loans instead of holding its assets in the form of government bonds, which expands the money supply. As the new loans are deposited in banks throughout the economy, these banks will,

in turn, loan out some of the deposits they receive, triggering the money multiplier discussed in [Money and Banking](#).

Assets		Liabilities + Net Worth	
Reserves	40	Deposits	400
Bonds	120		
Loans	300	Net Worth	60

(a) The original balance sheet

Assets		Liabilities + Net Worth	
Reserves	$40 + 20 = 60$	Deposits	400
Bonds	$120 - 20 = 100$		
Loans	300	Net Worth	60

(b) The central bank buys bonds

Assets		Liabilities + Net Worth	
Reserves	$60 - 20 = 40$	Deposits	400
Bonds	100		
Loans	$300 + 20 = 320$	Net Worth	60

(c) The bank makes additional loans

Where did the Federal Reserve get the \$20 million that it used to purchase the bonds? A central bank has the power to create money. In practical terms, the Federal Reserve would write a check to Happy Bank, so that Happy Bank can have that money credited to its bank account at the Federal Reserve. In truth, the Federal Reserve created the money to purchase the bonds out of thin air—or with a few clicks on some computer keys.

Open market operations can also reduce the quantity of money and loans in an economy. [\[link\]](#) (a) shows the balance sheet of Happy Bank before the central bank sells bonds in the open market. When Happy Bank purchases \$30 million in bonds, Happy Bank sends \$30 million of its reserves to the central bank, but now holds an additional \$30 million in bonds, as shown in [\[link\]](#) (b). However, Happy Bank wants to hold \$40 million in reserves, as in [\[link\]](#) (a), so it will adjust down the quantity of its loans by \$30 million, to bring its reserves back to the desired level, as shown in [\[link\]](#) (c). In practical terms, a bank can easily reduce its quantity of loans. At any given time, a bank is receiving payments on loans that it made previously and also

making new loans. If the bank just slows down or briefly halts making new loans, and instead adds those funds to its reserves, then its overall quantity of loans will decrease. A decrease in the quantity of loans also means fewer deposits in other banks, and other banks reducing their lending as well, as the money multiplier discussed in [Money and Banking](#) takes effect. And what about all those bonds? How do they affect the money supply? Read the following Clear It Up feature for the answer.

Assets		Liabilities + Net Worth	
Reserves	40	Deposits	400
Bonds	120		
Loans	300	Net Worth	60

(a) The original balance sheet

Assets		Liabilities + Net Worth	
Reserves	$40 - 30 = 10$	Deposits	400
Bonds	$120 + 30 = 150$		
Loans	300	Net Worth	60

(b) The central bank sells bonds to the bank

Assets		Liabilities + Net Worth	
Reserves	$10 + 30 = 40$	Deposits	400
Bonds	150		
Loans	$300 - 30 = 270$	Net Worth	60

(c) The bank makes fewer loans

Note:

Does selling or buying bonds increase the money supply?

Is it a sale of bonds by the central bank which increases bank reserves and lowers interest rates or is it a purchase of bonds by the central bank? The easy way to keep track of this is to treat the central bank as being *outside* the banking system. When a central bank buys bonds, money is flowing from the central bank to individual banks in the economy, increasing the supply of money in circulation. When a central bank sells bonds, then money from individual banks in the economy is flowing into the central bank—reducing the quantity of money in the economy.

Changing Reserve Requirements

A second method of conducting monetary policy is for the central bank to raise or lower the **reserve requirement**, which, as we noted earlier, is the percentage of each bank's deposits that it is legally required to hold either as cash in their vault or on deposit with the central bank. If banks are required to hold a greater amount in reserves, they have less money available to lend out. If banks are allowed to hold a smaller amount in reserves, they will have a greater amount of money available to lend out.

In early 2015, the Federal Reserve required banks to hold reserves equal to 0% of the first \$14.5 million in deposits, then to hold reserves equal to 3% of the deposits up to \$103.6 million, and 10% of any amount above \$103.6 million. Small changes in the reserve requirements are made almost every year. For example, the \$103.6 million dividing line is sometimes bumped up or down by a few million dollars. In practice, large changes in reserve requirements are rarely used to execute monetary policy. A sudden demand that all banks increase their reserves would be extremely disruptive and difficult to comply with, while loosening requirements too much would create a danger of banks being unable to meet the demand for withdrawals.

Changing the Discount Rate

The Federal Reserve was founded in the aftermath of the Financial Panic of 1907 when many banks failed as a result of bank runs. As mentioned earlier, since banks make profits by lending out their deposits, no bank, even those that are not bankrupt, can withstand a bank run. As a result of the Panic, the Federal Reserve was founded to be the “lender of last resort.” In the event of a bank run, sound banks, (banks that were not bankrupt) could borrow as much cash as they needed from the Fed's discount “window” to quell the bank run. The interest rate banks pay for such loans is called the **discount rate**. (They are so named because loans are made against the bank's outstanding loans “at a discount” of their face value.) Once depositors became convinced that the bank would be able to honor their withdrawals, they no longer had a reason to make a run on the bank. In short, the Federal Reserve was originally intended to provide credit

passively, but in the years since its founding, the Fed has taken on a more active role with monetary policy.

So, the third traditional method for conducting monetary policy is to raise or lower the discount rate. If the central bank raises the discount rate, then commercial banks will reduce their borrowing of reserves from the Fed, and instead call in loans to replace those reserves. Since fewer loans are available, the money supply falls and market interest rates rise. If the central bank lowers the discount rate it charges to banks, the process works in reverse.

In recent decades, the Federal Reserve has made relatively few discount loans. Before a bank borrows from the Federal Reserve to fill out its required reserves, the bank is expected to first borrow from other available sources, like other banks. This is encouraged by Fed's charging a higher discount rate, than the federal funds rate. Given that most banks borrow little at the discount rate, changing the discount rate up or down has little impact on their behavior. More importantly, the Fed has found from experience that open market operations are a more precise and powerful means of executing any desired monetary policy.

In the Federal Reserve Act, the phrase "...to afford means of rediscounting commercial paper" is contained in its long title. This tool was seen as the main tool for monetary policy when the Fed was initially created. This illustrates how monetary policy has evolved and how it continues to do so.

Key Concepts and Summary

A central bank has three traditional tools to conduct monetary policy: open market operations, which involves buying and selling government bonds with banks; reserve requirements, which determine what level of reserves a bank is legally required to hold; and discount rates, which is the interest rate charged by the central bank on the loans that it gives to other commercial banks. The most commonly used tool is open market operations.

Self-Check Questions

Exercise:**Problem:**

If the central bank sells \$500 in bonds to a bank that has issued \$10,000 in loans and is exactly meeting the reserve requirement of 10%, what will happen to the amount of loans and to the money supply in general?

Solution:

The bank has to hold \$1,000 in reserves, so when it buys the \$500 in bonds, it will have to reduce its loans by \$500 to make up the difference. The money supply decreases by the same amount.

Exercise:**Problem:**

What would be the effect of increasing the reserve requirements of banks on the money supply?

Solution:

An increase in reserve requirements would reduce the supply of money, since more money would be held in banks rather than circulating in the economy.

Review Questions**Exercise:****Problem:**

Explain how to use an open market operation to expand the money supply.

Exercise:

Problem:

Explain how to use the reserve requirement to expand the money supply.

Exercise:**Problem:**

Explain how to use the discount rate to expand the money supply.

Critical Thinking Question**Exercise:****Problem:**

Explain what would happen if banks were notified they had to increase their required reserves by one percentage point from, say, 9% to 10% of deposits. What would their options be to come up with the cash?

Problems**Exercise:****Problem:**

Suppose the Fed conducts an open market purchase by buying \$10 million in Treasury bonds from Acme Bank. Sketch out the balance sheet changes that will occur as Acme converts the bond sale proceeds to new loans. The initial Acme bank balance sheet contains the following information: Assets – reserves 30, bonds 50, and loans 50; Liabilities – deposits 300 and equity 30.

Exercise:

Problem:

Suppose the Fed conducts an open market sale by selling \$10 million in Treasury bonds to Acme Bank. Sketch out the balance sheet changes that will occur as Acme restores its required reserves (10% of deposits) by reducing its loans. The initial balance sheet for Acme Bank contains the following information: Assets – reserves 30, bonds 50, and loans 250; Liabilities – deposits 300 and equity 30.

References

Board of Governors of the Federal Reserve System. "Federal Open Market Committee." Accessed September 3, 2013.

<http://www.federalreserve.gov/monetarypolicy/fomc.htm>.

Board of Governors of the Federal Reserve System. "Reserve Requirements." Accessed November 5, 2013.

<http://www.federalreserve.gov/monetarypolicy/reservereq.htm>.

Cox, Jeff. 2014. "Fed Completes the Taper." Accessed March 31, 2015.

<http://www.cnn.com/id/102132961>.

Jahan, Sarwat. n.d. "Inflation Targeting: Holding the Line." International Monetary Fund. Accessed March 31, 2015.

<http://www.imf.org/external/pubs/ft/fandd/basics/target.htm>.

Glossary

discount rate

the interest rate charged by the central bank on the loans that it gives to other commercial banks

open market operations

the central bank selling or buying Treasury bonds to influence the quantity of money and the level of interest rates

reserve requirement

the percentage amount of its total deposits that a bank is legally obligated to either hold as cash in their vault or deposit with the central bank

Monetary Policy and Economic Outcomes

By the end of this section, you will be able to:

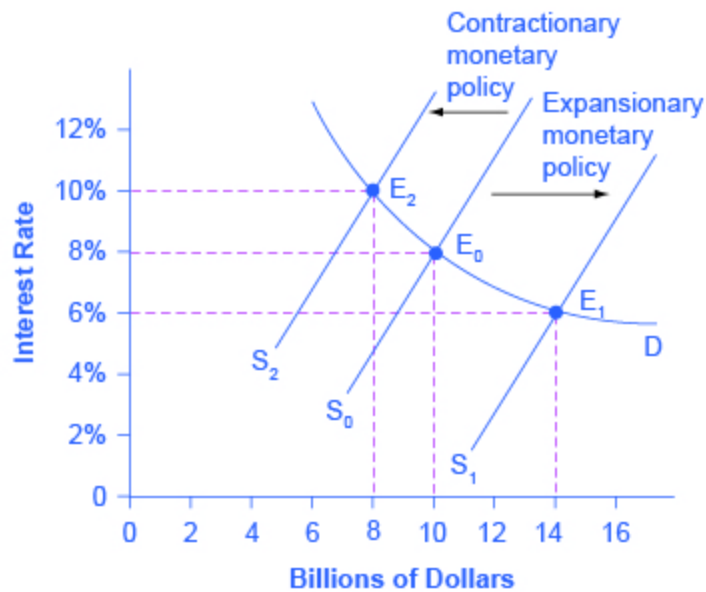
- Contrast expansionary monetary policy and contractionary monetary policy
- Explain how monetary policy impacts interest rates and aggregate demand
- Evaluate Federal Reserve decisions over the last forty years
- Explain the significance of quantitative easing (QE)

A monetary policy that lowers interest rates and stimulates borrowing is known as an **expansionary monetary policy** or **loose monetary policy**. Conversely, a monetary policy that raises interest rates and reduces borrowing in the economy is a **contractionary monetary policy** or **tight monetary policy**. This module will discuss how expansionary and contractionary monetary policies affect interest rates and aggregate demand, and how such policies will affect macroeconomic goals like unemployment and inflation. We will conclude with a look at the Fed's monetary policy practice in recent decades.

The Effect of Monetary Policy on Interest Rates

Consider the market for loanable bank funds, shown in [\[link\]](#). The original equilibrium (E_0) occurs at an interest rate of 8% and a quantity of funds loaned and borrowed of \$10 billion. An expansionary monetary policy will shift the supply of loanable funds to the right from the original supply curve (S_0) to S_1 , leading to an equilibrium (E_1) with a lower interest rate of 6% and a quantity of funds loaned of \$14 billion. Conversely, a contractionary monetary policy will shift the supply of loanable funds to the left from the original supply curve (S_0) to S_2 , leading to an equilibrium (E_2) with a higher interest rate of 10% and a quantity of funds loaned of \$8 billion.

Monetary Policy and Interest Rates



The original equilibrium occurs at E_0 . An expansionary monetary policy will shift the supply of loanable funds to the right from the original supply curve (S_0) to the new supply curve (S_1) and to a new equilibrium of E_1 , reducing the interest rate from 8% to 6%. A contractionary monetary policy will shift the supply of loanable funds to the left from the original supply curve (S_0) to the new supply (S_2), and raise the interest rate from 8% to 10%.

So how does a central bank “raise” interest rates? When describing the monetary policy actions taken by a central bank, it is common to hear that the central bank “raised interest rates” or “lowered interest rates.” We need to be clear about this: more precisely, through open market operations the central bank changes bank reserves in a way which affects the supply curve of loanable funds. As a result, interest rates change, as shown in [\[link\]](#). If they do not meet the Fed’s target, the Fed can supply more or less reserves until interest rates do.

Recall that the specific interest rate the Fed targets is the **federal funds rate**. The Federal Reserve has, since 1995, established its target federal funds rate in advance of any open market operations.

Of course, financial markets display a wide range of interest rates, representing borrowers with different risk premiums and loans that are to be repaid over different periods of time. In general, when the federal funds rate drops substantially, other interest rates drop, too, and when the federal funds rate rises, other interest rates rise. However, a fall or rise of one percentage point in the federal funds rate—which remember is for borrowing overnight—will typically have an effect of less than one percentage point on a 30-year loan to purchase a house or a three-year loan to purchase a car. Monetary policy can push the entire spectrum of interest rates higher or lower, but the specific interest rates are set by the forces of supply and demand in those specific markets for lending and borrowing.

The Effect of Monetary Policy on Aggregate Demand

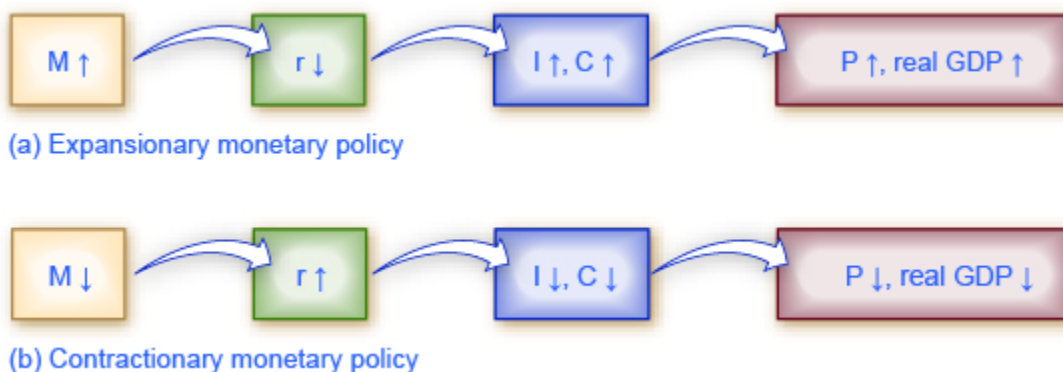
Monetary policy affects interest rates and the available quantity of loanable funds, which in turn affects several components of aggregate demand. Tight or contractionary monetary policy that leads to higher interest rates and a reduced quantity of loanable funds will reduce two components of aggregate demand. Business investment will decline because it is less attractive for firms to borrow money, and even firms that have money will notice that, with higher interest rates, it is relatively more attractive to put those funds in a financial investment than to make an investment in physical capital. In addition, higher interest rates will discourage consumer borrowing for big-ticket items like houses and cars. Conversely, loose or expansionary monetary policy that leads to lower interest rates and a higher quantity of loanable funds will tend to increase business investment and consumer borrowing for big-ticket items.

If the economy is suffering a recession and high unemployment, with output below potential GDP, expansionary monetary policy can help the economy return to potential GDP. An expansionary monetary policy will reduce interest rates and stimulate investment and consumption spending.

Conversely, if an economy is producing at a quantity of output above its potential GDP, a contractionary monetary policy can reduce the inflationary pressures for a rising price level. A contractionary monetary policy will raise interest rates, discourage borrowing for investment and consumption spending.

These examples suggest that monetary policy should be **countercyclical**; that is, it should act to counterbalance the business cycles of economic downturns and upswings. Monetary policy should be loosened when a recession has caused unemployment to increase and tightened when inflation threatens. Of course, countercyclical policy does pose a danger of overreaction. If loose monetary policy seeking to end a recession goes too far, it may push aggregate demand so far to the right that it triggers inflation. If tight monetary policy seeking to reduce inflation goes too far, it may push aggregate demand so far to the left that a recession begins. [\[link\]](#) (a) summarizes the chain of effects that connect loose and tight monetary policy to changes in output and the price level.

The Pathways of Monetary Policy



(a) In expansionary monetary policy the central bank causes the supply of money and loanable funds to increase, which lowers the interest rate, stimulating additional borrowing for investment and consumption, and shifting aggregate demand right. The result is a higher price level and, at least in the short run, higher real GDP. (b)

In contractionary monetary policy, the central bank causes the supply of money and credit in the economy to decrease, which raises the interest rate, discouraging borrowing for investment and consumption, and shifting aggregate demand left. The result is a lower price level and, at least in the short run, lower real GDP.

Quantitative Easing

The most powerful and commonly used of the three traditional tools of monetary policy—open market operations—works by expanding or contracting the money supply in a way that influences the interest rate. In late 2008, as the U.S. economy struggled with recession, the Federal Reserve had already reduced the interest rate to near-zero. With the recession still ongoing, the Fed decided to adopt an innovative and nontraditional policy known as **quantitative easing (QE)**. This is the purchase of long-term government and private mortgage-backed securities by central banks to make credit available so as to stimulate aggregate demand.

Quantitative easing differed from traditional monetary policy in several key ways. First, it involved the Fed purchasing long term Treasury bonds, rather than short term Treasury bills. In 2008, however, it was impossible to stimulate the economy any further by lowering short term rates because they were already as low as they could get. (Read the closing Bring it Home feature for more on this.) Therefore, Bernanke sought to lower long-term rates utilizing quantitative easing.

This leads to a second way QE is different from traditional monetary policy. Instead of purchasing Treasury securities, the Fed also began purchasing private mortgage-backed securities, something it had never done before. During the financial crisis, which precipitated the recession, mortgage-backed securities were termed “toxic assets,” because when the housing market collapsed, no one knew what these securities were worth, which put the financial institutions which were holding those securities on very shaky ground. By offering to purchase mortgage-backed securities, the Fed was both pushing long term interest rates down and also removing possibly “toxic assets” from the balance sheets of private financial firms, which would strengthen the financial system.

Quantitative easing (QE) occurred in three episodes:

1. During QE₁, which began in November 2008, the Fed purchased \$600 billion in mortgage-backed securities from government enterprises Fannie Mae and Freddie Mac.
2. In November 2010, the Fed began QE₂, in which it purchased \$600 billion in U.S. Treasury bonds.
3. QE₃, began in September 2012 when the Fed commenced purchasing \$40 billion of additional mortgage-backed securities per month. This amount was increased in December 2012 to \$85 billion per month. The Fed stated that, when economic conditions permit, it will begin tapering (or reducing the monthly purchases). By October 2014, the Fed had announced the final \$15 billion purchase of bonds, ending Quantitative Easing.

The quantitative easing policies adopted by the Federal Reserve (and by other central banks around the world) are usually thought of as temporary emergency measures. If these steps are, indeed, to be temporary, then the Federal Reserve will need to stop making these additional loans and sell off the financial securities it has accumulated. The concern is that the process of quantitative easing may prove more difficult to reverse than it was to enact. The evidence suggests that QE₁ was somewhat successful, but that QE₂ and QE₃ have been less so.

Key Concepts and Summary

An expansionary (or loose) monetary policy raises the quantity of money and credit above what it otherwise would have been and reduces interest rates, boosting aggregate demand, and thus countering recession. A contractionary monetary policy, also called a tight monetary policy, reduces the quantity of money and credit below what it otherwise would have been and raises interest rates, seeking to hold down inflation. During the 2008–2009 recession, central banks around the world also used quantitative easing to expand the supply of credit.

Self-Check Questions

Exercise:

Problem:

Why does contractionary monetary policy cause interest rates to rise?

Solution:

Contractionary policy reduces the amount of loanable funds in the economy. As with all goods, greater scarcity leads a greater price, so the interest rate, or the price of borrowing money, rises.

Exercise:**Problem:**

Why does expansionary monetary policy causes interest rates to drop?

Solution:

An increase in the amount of available loanable funds means that there are more people who want to lend. They, therefore, bid the price of borrowing (the interest rate) down.

Review Questions**Exercise:****Problem:**

How do the expansionary and contractionary monetary policy affect the quantity of money?

Exercise:**Problem:**

How do tight and loose monetary policy affect interest rates?

Exercise:

Problem:

How do expansionary, tight, contractionary, and loose monetary policy affect aggregate demand?

Exercise:**Problem:**

Which kind of monetary policy would you expect in response to high inflation: expansionary or contractionary? Why?

Exercise:**Problem:**

Explain how to use quantitative easing to stimulate aggregate demand.

Critical Thinking Question**Exercise:****Problem:**

A well-known economic model called the Phillips Curve (discussed in [The Keynesian Perspective](#) chapter) describes the short run tradeoff typically observed between inflation and unemployment. Based on the discussion of expansionary and contractionary monetary policy, explain why one of these variables usually falls when the other rises.

Glossary

contractionary monetary policy

a monetary policy that reduces the supply of money and loans

countercyclical

moving in the opposite direction of the business cycle of economic downturns and upswings

expansionary monetary policy

a monetary policy that increases the supply of money and the quantity of loans

federal funds rate

the interest rate at which one bank lends funds to another bank overnight

loose monetary policy

see expansionary monetary policy

quantitative easing (QE)

the purchase of long term government and private mortgage-backed securities by central banks to make credit available in hopes of stimulating aggregate demand

tight monetary policy

see contractionary monetary policy

Pitfalls for Monetary Policy

By the end of this section, you will be able to:

- Analyze whether monetary policy decisions should be made more democratically
- Calculate the velocity of money
- Evaluate the central bank's influence on inflation, unemployment, asset bubbles, and leverage cycles
- Calculate the effects of monetary stimulus

In the real world, effective monetary policy faces a number of significant hurdles. Monetary policy affects the economy only after a time lag that is typically long and of variable length. Remember, monetary policy involves a chain of events: the central bank must perceive a situation in the economy, hold a meeting, and make a decision to react by tightening or loosening monetary policy. The change in monetary policy must percolate through the banking system, changing the quantity of loans and affecting interest rates. When interest rates change, businesses must change their investment levels and consumers must change their borrowing patterns when purchasing homes or cars. Then it takes time for these changes to filter through the rest of the economy.

As a result of this chain of events, monetary policy has little effect in the immediate future; instead, its primary effects are felt perhaps one to three years in the future. The reality of long and variable time lags does not mean that a central bank should refuse to make decisions. It does mean that central banks should be humble about taking action, because of the risk that their actions can create as much or more economic instability as they resolve.

Excess Reserves

Banks are legally required to hold a minimum level of reserves, but no rule prohibits them from holding additional **excess reserves** above the legally mandated limit. For example, during a recession banks may be hesitant to lend, because they fear that when the economy is contracting, a high proportion of loan applicants become less likely to repay their loans.

When many banks are choosing to hold excess reserves, expansionary monetary policy may not work well. This may occur because the banks are concerned about a deteriorating economy, while the central bank is trying to expand the money supply. If the banks prefer to hold excess reserves above the legally required level, the central bank cannot force individual banks to make loans. Similarly, sensible businesses and consumers may be reluctant to borrow substantial amounts of money in a recession, because they recognize that firms' sales and employees' jobs are more insecure in a recession, and they do not want to face the need to make interest payments. The result is that during an especially deep recession, an expansionary monetary policy may have little effect on either the price level or the real GDP.

Japan experienced this situation in the 1990s and early 2000s. Japan's economy entered a period of very slow growth, dipping in and out of recession, in the early 1990s. By February 1999, the Bank of Japan had lowered the equivalent of its federal funds rate to 0%. It kept it there most of the time through 2003. Moreover, in the two years from March 2001 to March 2003, the Bank of Japan also expanded the money supply of the country by about 50%—an enormous increase. Even this highly expansionary monetary policy, however, had no substantial effect on stimulating aggregate demand. Japan's economy continued to experience extremely slow growth into the mid-2000s.

Note:

Should monetary policy decisions be made more democratically?

Should monetary policy be conducted by a nation's Congress or legislature comprised of elected representatives? Or should it be conducted by a politically appointed central bank that is more independent of voters? Here are some of the arguments made by each side.

The Case for Greater Democratic Control of Monetary Policy

Elected representatives conduct fiscal policy by passing tax and spending bills. They could handle monetary policy in the same way. Sure, they will sometimes make mistakes, but in a democracy, it is better to have mistakes made by elected officials accountable to voters than by political appointees. After all, the people appointed to the top governing positions at the Federal Reserve—and to most central banks around the world—are typically bankers

and economists. They are not representatives of borrowers like small businesses or farmers nor are they representatives of labor unions. Central banks might not be so quick to raise interest rates if they had to pay more attention to firms and people in the real economy.

The Case for an Independent Central Bank

Because the central bank has some insulation from day-to-day politics, its members can take a nonpartisan look at specific economic situations and make tough, immediate decisions when necessary. The idea of giving a legislature the ability to create money and hand out loans is likely to end up badly, sooner or later. It is simply too tempting for lawmakers to expand the money supply to fund their projects. The long term result will be rampant inflation. Also, a central bank, acting according to the laws passed by elected officials, can respond far more quickly than a legislature. For example, the U.S. budget takes months to debate, pass, and be signed into law, but monetary policy decisions can be made much more rapidly. Day-to-day democratic control of monetary policy is impractical and seems likely to lead to an overly expansionary monetary policy and higher inflation.

The problem of excess reserves does not affect contractionary policy. Central bankers have an old saying that monetary policy can be like pulling and pushing on a string: when the central bank pulls on the string and uses contractionary monetary policy, it can definitely raise interest rates and reduce aggregate demand. However, when the central bank tries to push on the string of expansionary monetary policy, the string may sometimes just fold up limp and have little effect, because banks decide not to loan out their excess reserves. This analogy should not be taken too literally—expansionary monetary policy usually does have real effects, after that inconveniently long and variable lag. There are also times, like Japan's economy in the late 1990s and early 2000s, when expansionary monetary policy has been insufficient to lift a recession-prone economy.

Unpredictable Movements of Velocity

Velocity is a term that economists use to describe how quickly money circulates through the economy. The **velocity** of money in a year is defined as:

Equation:

$$\text{Velocity} = \frac{\text{nominal GDP}}{\text{money supply}}$$

Specific measurements of velocity depend on the definition of the money supply being used. Consider the velocity of M1, the total amount of currency in circulation and checking account balances. In 2009, for example, M1 was \$1.7 trillion and nominal GDP was \$14.3 trillion, so the velocity of M1 was 8.4 (\$14.3 trillion/\$1.7 trillion). A higher velocity of money means that the average dollar circulates more times in a year; a lower velocity means that the average dollar circulates fewer times in a year.

Perhaps you heard the “d” word mentioned during our recent economic downturn. See the following Clear It Up feature for a discussion of how deflation could affect monetary policy.

Note:

What happens during episodes of deflation?

Deflation occurs when the rate of inflation is negative; that is, instead of money having less purchasing power over time, as occurs with inflation, money is worth more. Deflation can make it very difficult for monetary policy to address a recession.

Remember that the real interest rate is the nominal interest rate minus the rate of inflation. If the nominal interest rate is 7% and the rate of inflation is 3%, then the borrower is effectively paying a 4% real interest rate. If the nominal interest rate is 7% and there is *deflation* of 2%, then the real interest rate is actually 9%. In this way, an unexpected deflation raises the real interest payments for borrowers. It can lead to a situation where an unexpectedly high number of loans are not repaid, and banks find that their net worth is decreasing or negative. When banks are suffering losses, they become less able and eager to make new loans. Aggregate demand declines, which can lead to recession.

Then the double-whammy: After causing a recession, deflation can make it difficult for monetary policy to work. Say that the central bank uses expansionary monetary policy to reduce the nominal interest rate all the way

to zero—but the economy has 5% deflation. As a result, the real interest rate is 5%, and because a central bank cannot make the nominal interest rate negative, expansionary policy cannot reduce the real interest rate further. In the U.S. economy during the early 1930s, deflation was 6.7% per year from 1930–1933, which caused many borrowers to default on their loans and many banks to end up bankrupt, which in turn contributed substantially to the Great Depression. Not all episodes of deflation, however, end in economic depression. Japan, for example, experienced deflation of slightly less than 1% per year from 1999–2002, which hurt the Japanese economy, but it still grew by about 0.9% per year over this period. Indeed, there is at least one historical example of deflation coexisting with rapid growth. The U.S. economy experienced deflation of about 1.1% per year over the quarter-century from 1876–1900, but real GDP also expanded at a rapid clip of 4% per year over this time, despite some occasional severe recessions. The central bank should be on guard against deflation and, if necessary, use expansionary monetary policy to prevent any long-lasting or extreme deflation from occurring. Except in severe cases like the Great Depression, deflation does not guarantee economic disaster.

Changes in velocity can cause problems for monetary policy. To understand why, rewrite the definition of velocity so that the money supply is on the left-hand side of the equation. That is:

Equation:

$$\text{Money supply} \times \text{velocity} = \text{Nominal GDP}$$

Recall from [The Macroeconomic Perspective](#) that

Equation:

$$\text{Nominal GDP} = \text{Price Level (or GDP Deflator)} \times \text{Real GDP}.$$

Therefore,

Equation:

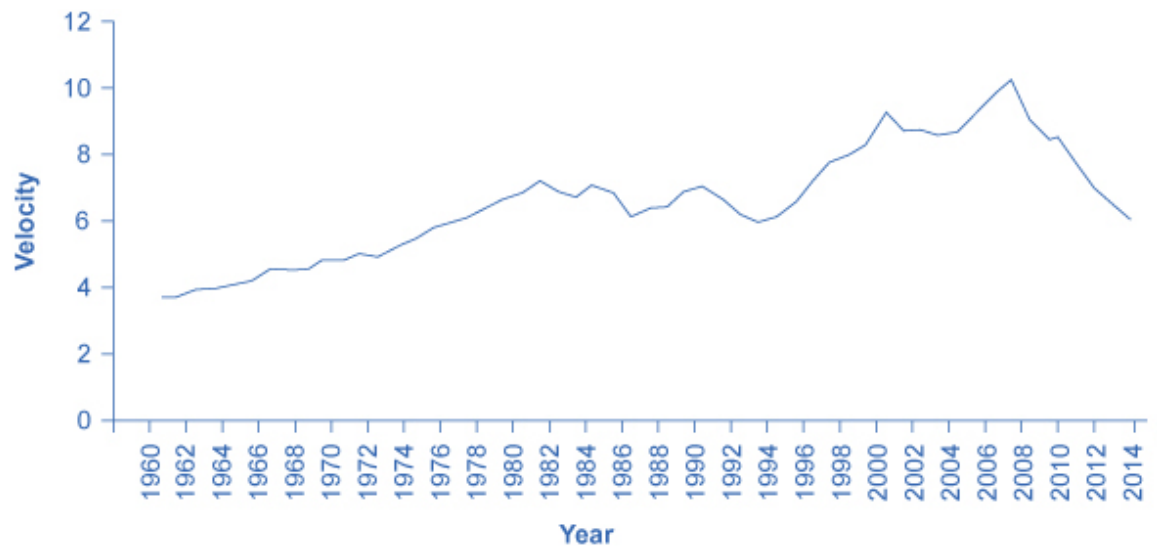
$$\text{Money Supply} \times \text{velocity} = \text{Nominal GDP} = \text{Price Level} \times \text{Real GDP}.$$

This equation is sometimes called the **basic quantity equation of money** but, as you can see, it is just the definition of velocity written in a different form. This equation must hold true, by definition.

If velocity is constant over time, then a certain percentage rise in the money supply on the left-hand side of the basic quantity equation of money will inevitably lead to the same percentage rise in nominal GDP—although this change could happen through an increase in inflation, or an increase in real GDP, or some combination of the two. If velocity is changing over time but in a constant and predictable way, then changes in the money supply will continue to have a predictable effect on nominal GDP. If velocity changes unpredictably over time, however, then the effect of changes in the money supply on nominal GDP becomes unpredictable.

The actual velocity of money in the U.S. economy as measured by using M1, the most common definition of the money supply, is illustrated in [\[link\]](#). From 1960 up to about 1980, velocity appears fairly predictable; that is, it is increasing at a fairly constant rate. In the early 1980s, however, velocity as calculated with M1 becomes more variable. The reasons for these sharp changes in velocity remain a puzzle. Economists suspect that the changes in velocity are related to innovations in banking and finance which have changed how money is used in making economic transactions: for example, the growth of electronic payments; a rise in personal borrowing and credit card usage; and accounts that make it easier for people to hold money in savings accounts, where it is counted as M2, right up to the moment that they want to write a check on the money and transfer it to M1. So far at least, it has proven difficult to draw clear links between these kinds of factors and the specific up-and-down fluctuations in M1. Given many changes in banking and the prevalence of electronic banking, M2 is now favored as a measure of money rather than the narrower M1.

Velocity Calculated Using M1



Velocity is the nominal GDP divided by the money supply for a given year. Different measures of velocity can be calculated by using different measures of the money supply. Velocity, as calculated by using M1, has lacked a steady trend since the 1980s, instead bouncing up and down. (credit: Federal Reserve Bank of St. Louis)

In the 1970s, when velocity as measured by M1 seemed predictable, a number of economists, led by Nobel laureate Milton Friedman (1912–2006), argued that the best monetary policy was for the central bank to increase the money supply at a constant growth rate. These economists argued that with the long and variable lags of monetary policy, and the political pressures on central bankers, central bank monetary policies were as likely to have undesirable as to have desirable effects. Thus, these economists believed that the monetary policy should seek steady growth in the money supply of 3% per year. They argued that a steady rate of monetary growth would be correct over longer time periods, since it would roughly match the growth of the real economy. In addition, they argued that giving the central bank less discretion to conduct monetary policy would prevent an overly activist central bank from becoming a source of economic instability and uncertainty. In this spirit, Friedman wrote in 1967: “The first and most important lesson that history teaches about what monetary policy can do—and it is a lesson of the most

profound importance—is that monetary policy can prevent money itself from being a major source of economic disturbance.”

As the velocity of M1 began to fluctuate in the 1980s, having the money supply grow at a predetermined and unchanging rate seemed less desirable, because as the quantity theory of money shows, the combination of constant growth in the money supply and fluctuating velocity would cause nominal GDP to rise and fall in unpredictable ways. The jumpiness of velocity in the 1980s caused many central banks to focus less on the rate at which the quantity of money in the economy was increasing, and instead to set monetary policy by reacting to whether the economy was experiencing or in danger of higher inflation or unemployment.

Unemployment and Inflation

If you were to survey central bankers around the world and ask them what they believe should be the primary task of monetary policy, the most popular answer by far would be fighting inflation. Most central bankers believe that the neoclassical model of economics accurately represents the economy over the medium to long term which is that an expansionary monetary policy only creates an inflationary increase in the price level, but it does not alter GDP or unemployment. From this perspective, all that monetary policy can do is to lead to low inflation or high inflation—and low inflation provides a better climate for a healthy and growing economy. After all, low inflation means that businesses making investments can focus on real economic issues, not on figuring out ways to protect themselves from the costs and risks of inflation. In this way, a consistent pattern of low inflation can contribute to long-term growth.

This vision of focusing monetary policy on a low rate of inflation is so attractive that many countries have rewritten their central banking laws since in the 1990s to have their bank practice **inflation targeting**, which means that the central bank is legally required to focus primarily on keeping inflation low. By 2014, central banks in 28 countries, including Austria, Brazil, Canada, Israel, Korea, Mexico, New Zealand, Spain, Sweden, Thailand, and the United Kingdom faced a legal requirement to target the inflation rate. A notable exception is the Federal Reserve in the United States, which does not

practice inflation-targeting. Instead, the law governing the Federal Reserve requires it to take both unemployment and inflation into account.

Economists have no final consensus on whether a central bank should be required to focus only on inflation or should have greater discretion. For those who subscribe to the inflation targeting philosophy, the fear is that politicians who are worried about slow economic growth and unemployment will constantly pressure the central bank to conduct a loose monetary policy—even if the economy is already producing at potential GDP. In some countries, the central bank may lack the political power to resist such pressures, with the result of higher inflation, but no long-term reduction in unemployment. The U.S. Federal Reserve has a tradition of independence, but central banks in other countries may be under greater political pressure. For all of these reasons—long and variable lags, excess reserves, unstable velocity, and controversy over economic goals—monetary policy in the real world is often difficult. The basic message remains, however, that central banks can affect aggregate demand through the conduct of monetary policy and in that way influence macroeconomic outcomes.

Monetary Stimulus

Let's end this chapter with a Work it Out exercise in how the Fed—or any central bank—would stir up the economy by increasing the money supply.

Note:

Calculating the Effects of Monetary Stimulus

Suppose that the central bank wants to stimulate the economy by increasing the money supply. The bankers estimate that the velocity of money is 3, and that the price level will increase from 100 to 110 due to the stimulus. Using the quantity equation of money, what will be the impact of an \$800 billion dollar increase in the money supply on the quantity of goods and services in the economy given an initial money supply of \$4 trillion?

Step 1. We begin by writing the quantity equation of money: $MV = PQ$. We know that initially $V = 3$, $M = 4,000$ (billion) and $P = 100$. Substituting these numbers in, we can solve for Q :

Equation:

$$\begin{aligned}MV &= PQ \\4,000 \times 3 &= 100 \times Q \\Q &= 120\end{aligned}$$

Step 2. Now we want to find the effect of the addition \$800 billion in the money supply, together with the increase in the price level. The new equation is:

Equation:

$$\begin{aligned}MV &= PQ \\4,800 \times 3 &= 110 \times Q \\Q &= 130.9\end{aligned}$$

Step 3. If we take the difference between the two quantities, we find that the monetary stimulus increased the quantity of goods and services in the economy by 10.9 billion.

The discussion in this chapter has focused on domestic monetary policy; that is, the view of monetary policy within an economy. [Exchange Rates and International Capital Flows](#) explores the international dimension of monetary policy, and how monetary policy becomes involved with exchange rates and international flows of financial capital.

Note:

The Problem of the Zero Percent Interest Rate Lower Bound

In 2008, the U.S. Federal Reserve found itself in a difficult position. The federal funds rate was on its way to near zero, which meant that traditional open market operations, by which the Fed purchases U.S. Treasury Bills to lower short term interest rates, was no longer viable. This so called “zero bound problem,” prompted the Fed, under then Chair Ben Bernanke, to attempt some unconventional policies, collectively called quantitative easing. By early 2014, quantitative easing nearly quintupled the amount of bank reserves. This likely contributed to the U.S. economy’s recovery, but the impact was muted, probably due to some of the hurdles mentioned in the last

section of this module. The unprecedented increase in bank reserves also led to fears of inflation. As of early 2015, however, there have been no serious signs of a boom, with core inflation around a stable 1.7%.

Key Concepts and Summary

Monetary policy is inevitably imprecise, for a number of reasons: (a) the effects occur only after long and variable lags; (b) if banks decide to hold excess reserves, monetary policy cannot force them to lend; and (c) velocity may shift in unpredictable ways. The basic quantity equation of money is $MV = PQ$, where M is the money supply, V is the velocity of money, P is the price level, and Q is the real output of the economy. Some central banks, like the European Central Bank, practice inflation targeting, which means that the only goal of the central bank is to keep inflation within a low target range. Other central banks, such as the U.S. Federal Reserve, are free to focus on either reducing inflation or stimulating an economy that is in recession, whichever goal seems most important at the time.

Self-Check Questions

Exercise:

Problem:

Why might banks want to hold excess reserves in time of recession?

Solution:

In times of economic uncertainty, banks may worry that borrowers will lose the ability to repay their loans. They may also fear that a panic is more likely and they will need the excess reserves to meet their obligations.

Exercise:

Problem: Why might the velocity of money change unexpectedly?

Solution:

If consumer optimism changes, spending can speed up or slow down. This could also happen in a case where consumers need to buy a large number of items quickly, such as in a situation of national emergency.

Review Questions**Exercise:****Problem:**

Which kind of monetary policy would you expect in response to recession: expansionary or contractionary? Why?

Exercise:**Problem:**

How might each of the following factors complicate the implementation of monetary policy: long and variable lags, excess reserves, and movements in velocity?

Exercise:

Problem: Define the velocity of the money supply.

Exercise:

Problem: What is the basic quantity equation of money?

Exercise:

Problem: How does a monetary policy of inflation targeting work?

Critical Thinking Questions**Exercise:**

Problem:

How does rule-based monetary policy differ from discretionary monetary policy (that is, monetary policy not based on a rule)? What are some of the arguments for each?

Exercise:**Problem:**

Is it preferable for central banks to primarily target inflation or unemployment? Why?

Problems**Exercise:****Problem:**

All other things being equal, by how much will nominal GDP expand if the central bank increases the money supply by \$100 billion, and the velocity of money is 3? (Use this information as necessary to answer the following 4 questions.)

Exercise:**Problem:**

Suppose now that economists expect the velocity of money to increase by 50% as a result of the monetary stimulus. What will be the total increase in nominal GDP?

Exercise:**Problem:**

If GDP is 1,500 and the money supply is 400, what is velocity?

Exercise:

Problem:

If GDP now rises to 1,600, but the money supply does not change, how has velocity changed?

Exercise:**Problem:**

If GDP now falls back to 1,500 and the money supply falls to 350, what is velocity?

References

Tobin, James. "The Concise Encyclopedia of Economics: Monetary Policy." *Library of Economics and Liberty*. Accessed November 2013.
<http://www.econlib.org/library/Enc/MonetaryPolicy.html>.

Federal Reserve Bank of New York. "The Founding of the Fed." Accessed November 2013.
http://www.newyorkfed.org/aboutthefed/history_article.html.

Glossary

basic quantity equation of money

$$\text{money supply} \times \text{velocity} = \text{nominal GDP}$$

excess reserves

reserves banks hold that exceed the legally mandated limit

inflation targeting

a rule that the central bank is required to focus only on keeping inflation low

velocity

the speed with which money circulates through the economy; calculated as the nominal GDP divided by the money supply

Introduction to the Keynesian Perspective

class="introduction"

Signs of a Recession

Home
foreclosures
were just
one of the
many signs
and
symptoms
of the recent
Great
Recession.
During that
time, many
businesses
closed and
many
people lost
their jobs.
(Credit:
modificatio
n of work
by Taber
Andrew
Bain/Flickr
Creative
Commons)



Note:

The Great Recession

The Great Recession of 2008–2009 hit the U.S. economy hard. According to the Bureau of Labor Statistics (BLS), the number of unemployed Americans rose from 6.8 million in May 2007 to 15.4 million in October 2009. During that time, the U.S. Census Bureau estimated that approximately 170,000 small businesses closed. Mass layoffs peaked in February 2009 when 326,392 workers were given notice. U.S. productivity and output fell as well. Job losses, declining home values, declining incomes, and uncertainty about the future caused consumption expenditures to decrease. According to the BLS, household spending dropped by 7.8%.

Home foreclosures and the meltdown in U.S. financial markets called for immediate action by Congress, the President, and the Federal Reserve Bank. For example, programs such as the American Restoration and Recovery Act were implemented to help millions of people by providing tax credits for homebuyers, paying “cash for clunkers,” and extending unemployment benefits. From cutting back on spending, filing for unemployment, and losing homes, millions of people were affected by the

recession. And while the United States is now on the path to recovery, the impact will be felt for many years to come.

What caused this recession and what prevented the economy from spiraling further into another depression? Policymakers looked to the lessons learned from the Great Depression of the 1930s and to the models developed by John Maynard Keynes to analyze the causes and find solutions to the country's economic woes. The Keynesian perspective is the subject of this chapter.

Note:

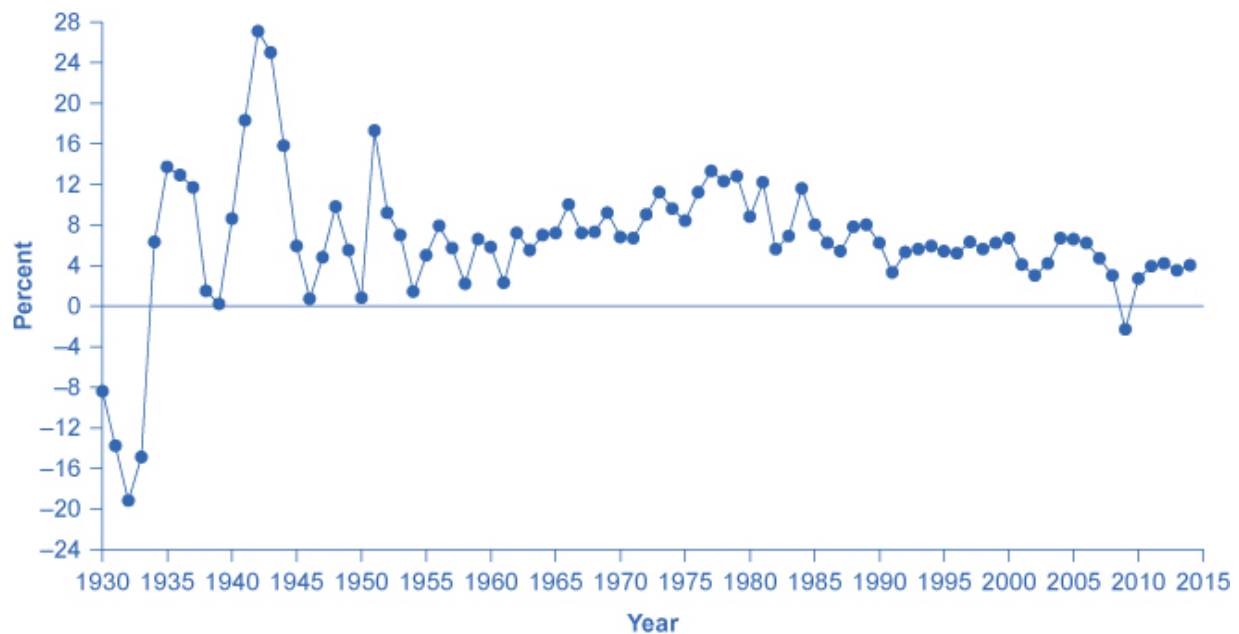
Introduction to the Keynesian Perspective

In this chapter, you will learn about:

- Aggregate Demand in Keynesian Analysis
- The Building Blocks of Keynesian Analysis
- The Phillips Curve
- The Keynesian Perspective on Market Forces

We have learned that the level of economic activity, for example output, employment, and spending, tends to grow over time. In [The Keynesian Perspective](#) we learned the reasons for this trend. [The Macroeconomic Perspective](#) pointed out that the economy tends to cycle around the long-run trend. In other words, the economy does not always grow at its average growth rate. Sometimes economic activity grows at the trend rate, sometimes it grows more than the trend, sometimes it grows less than the trend, and sometimes it actually declines. You can see this cyclical behavior in [\[link\]](#).

U.S. Gross Domestic Product, Percent Changes 1930–2014



The chart tracks the percent change in GDP since 1930. The magnitude of both recessions and peaks was quite large between 1930 and 1945. (Source: Bureau of Economic Analysis, “National Economic Accounts”)

This empirical reality raises two important questions: How can we explain the cycles, and to what extent can they be moderated? This chapter (on the Keynesian perspective) and [The Neoclassical Perspective](#) explore those questions from two different points of view, building on what we learned in [The Aggregate Demand/Aggregate Supply Model](#).

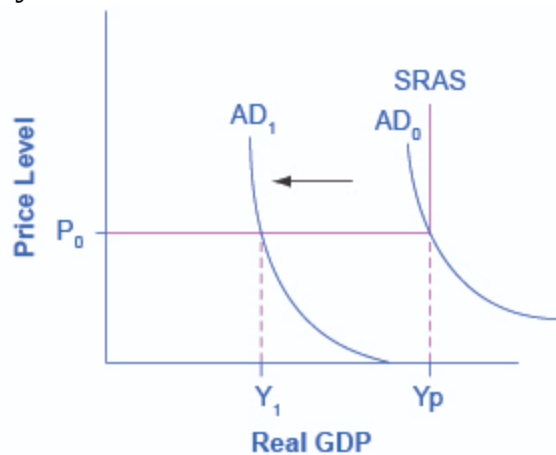
Aggregate Demand in Keynesian Analysis

By the end of this section, you will be able to:

- Explain real GDP, recessionary gaps, and inflationary gaps
- Recognize the Keynesian AD/AS model
- Identify the determining factors of both consumption expenditure and investment expenditure
- Analyze the factors that determine government spending and net exports

The Keynesian perspective focuses on aggregate demand. The idea is simple: firms produce output only if they expect it to sell. Thus, while the availability of the factors of production determines a nation's potential GDP, the amount of goods and services actually being sold, known as **real GDP**, depends on how much demand exists across the economy. This point is illustrated in [\[link\]](#).

The Keynesian AD/AS Model



The Keynesian View of the AD/AS Model uses an SRAS curve, which is horizontal at levels of output below potential and vertical at potential output. Thus, when beginning from potential output, any decrease in AD affects only output, but not prices; any increase in AD affects only prices, not output.

Keynes argued that, for reasons we explain shortly, aggregate demand is not stable—that it can change unexpectedly. Suppose the economy starts where AD intersects SRAS at P_0 and Y_p . Because Y_p is potential output, the economy is at full employment. Because AD is volatile, it can easily fall. Thus, even if we start at Y_p , if AD falls, then we find ourselves in what Keynes termed a **recessionary gap**. The economy is in equilibrium but with less than full employment, as shown at Y_1 in the [\[link\]](#). Keynes believed that the economy would tend to stay in a recessionary gap, with its attendant unemployment, for a significant period of time.

In the same way (though not shown in the figure), if AD increases, the economy could experience an **inflationary gap**, where demand is attempting to push the economy past potential output. As a consequence, the economy experiences inflation. The key policy implication for either situation is that government needs to step in and close the gap, increasing spending during recessions and decreasing spending during booms to return aggregate demand to match potential output.

Recall from [The Aggregate Supply-Aggregate Demand Model](#) that aggregate demand is total spending, economy-wide, on domestic goods and services. (Aggregate demand (AD) is actually what economists call total planned expenditure. Read the appendix on [The Expenditure-Output Model](#) for more on this.) You may also remember that aggregate demand is the sum of four components: consumption expenditure, investment expenditure, government spending, and spending on net exports (exports minus imports). In the following sections, we will examine each component through the Keynesian perspective.

What Determines Consumption Expenditure?

Consumption expenditure is spending by households and individuals on durable goods, nondurable goods, and services. Durable goods are things that last and provide value over time, such as automobiles. Nondurable goods are things like groceries—once you consume them, they are gone. Recall from [The Macroeconomic Perspective](#) that services are intangible things consumers buy, like healthcare or entertainment.

Keynes identified three factors that affect consumption:

- Disposable income: For most people, the single most powerful determinant of how much they consume is how much income they have in their take-home pay, also known as **disposable income**, which is income after taxes.
- Expected future income: Consumer expectations about future income also are important in determining consumption. If consumers feel optimistic about the future, they are more likely to spend and increase overall aggregate demand. News of recession and troubles in the economy will make them pull back on consumption.
- Wealth or credit: When households experience a rise in wealth, they may be willing to consume a higher share of their income and to save less. When the U.S. stock market rose dramatically in the late 1990s, for example, U.S. rates of saving declined, probably in part because people felt that their wealth had increased and there was less need to save. How do people spend beyond their income, when they perceive their wealth increasing? The answer is borrowing. On the other side, when the U.S. stock market declined about 40% from March 2008 to March 2009, people felt far greater uncertainty about their economic future, so rates of saving increased while consumption declined.

Finally, Keynes noted that a variety of other factors combine to determine how much people save and spend. If household preferences about saving shift in a way that encourages consumption rather than saving, then AD will shift out to the right.

Note:

Visit this [website](#) for more information about how the recession affected various groups of people.



What Determines Investment Expenditure?

Spending on new capital goods is called investment expenditure.

Investment falls into four categories: producer's durable equipment and software, new nonresidential structures (such as factories, offices, and retail locations), changes in inventories, and residential structures (such as single-family homes, townhouses, and apartment buildings). The first three types of investment are conducted by businesses, while the last is conducted by households.

Keynes's treatment of investment focuses on the key role of expectations about the future in influencing business decisions. When a business decides to make an investment in physical assets, like plants or equipment, or in intangible assets, like skills or a research and development project, that firm considers both the expected benefits of the investment (expectations of future profits) and the costs of the investment (interest rates).

- **Expectations of future profits:** The clearest driver of the benefits of an investment is expectations for future profits. When an economy is expected to grow, businesses perceive a growing market for their products. Their higher degree of business confidence will encourage new investment. For example, in the second half of the 1990s, U.S. investment levels surged from 18% of GDP in 1994 to 21% in 2000. However, when a recession started in 2001, U.S. investment levels quickly sank back to 18% of GDP by 2002.
- **Interest rates** also play a significant role in determining how much investment a firm will make. Just as individuals need to borrow money to purchase homes, so businesses need financing when they purchase

big ticket items. The cost of investment thus includes the interest rate. Even if the firm has the funds, the interest rate measures the opportunity cost of purchasing business capital. Lower interest rates stimulate investment spending and higher interest rates reduce it.

Many factors can affect the expected profitability on investment. For example, if the price of energy declines, then investments that use energy as an input will yield higher profits. If government offers special incentives for investment (for example, through the tax code), then investment will look more attractive; conversely, if government removes special investment incentives from the tax code, or increases other business taxes, then investment will look less attractive. As Keynes noted, business investment is the most variable of all the components of aggregate demand.

What Determines Government Spending?

The third component of aggregate demand is spending by federal, state, and local governments. Although the United States is usually thought of as a market economy, government still plays a significant role in the economy. As we discuss in [Environmental Protection and Negative Externalities](#) and [Positive Externalities and Public Goods](#), government provides important public services such as national defense, transportation infrastructure, and education.

Keynes recognized that the government budget offered a powerful tool for influencing aggregate demand. Not only could AD be stimulated by more government spending (or reduced by less government spending), but consumption and investment spending could be influenced by lowering or raising tax rates. Indeed, Keynes concluded that during extreme times like deep recessions, only the government had the power and resources to move aggregate demand.

What Determines Net Exports?

Recall that exports are products produced domestically and sold abroad while imports are products produced abroad but purchased domestically. Since aggregate demand is defined as spending on domestic goods and

services, export expenditures add to AD, while import expenditures subtract from AD.

Two sets of factors can cause shifts in export and import demand: changes in relative growth rates between countries and changes in relative prices between countries. The level of demand for a nation’s exports tends to be most heavily affected by what is happening in the economies of the countries that would be purchasing those exports. For example, if major importers of American-made products like Canada, Japan, and Germany have recessions, exports of U.S. products to those countries are likely to decline. Conversely, the quantity of a nation’s imports is directly affected by the amount of income in the domestic economy: more income will bring a higher level of imports.

Exports and imports can also be affected by relative prices of goods in domestic and international markets. If U.S. goods are relatively cheaper compared with goods made in other places, perhaps because a group of U.S. producers has mastered certain productivity breakthroughs, then U.S. exports are likely to rise. If U.S. goods become relatively more expensive, perhaps because a change in the exchange rate between the U.S. dollar and other currencies has pushed up the price of inputs to production in the United States, then exports from U.S. producers are likely to decline.

[\[link\]](#) summarizes the reasons given here for changes in aggregate demand.

Reasons for a Decrease in Aggregate Demand	Reasons for an Increase in Aggregate Demand
---	--

Reasons for a Decrease in Aggregate Demand	Reasons for an Increase in Aggregate Demand
<p><i>Consumption</i></p> <ul style="list-style-type: none"> • Rise in taxes • Fall in income • Rise in interest • Desire to save more • Decrease in wealth • Fall in future expected income 	<p><i>Consumption</i></p> <ul style="list-style-type: none"> • Decrease in taxes • Increase in income • Fall in interest rates • Desire to save less • Rise in wealth • Rise in future expected income
<p><i>Investment</i></p> <ul style="list-style-type: none"> • Fall in expected rate of return • Rise in interest rates • Drop in business confidence 	<p><i>Investment</i></p> <ul style="list-style-type: none"> • Rise in expected rate of return • Drop in interest rates • Rise in business confidence
<p><i>Government</i></p> <ul style="list-style-type: none"> • Reduction in government spending • Increase in taxes 	<p><i>Government</i></p> <ul style="list-style-type: none"> • Increase in government spending • Decrease in taxes
<p><i>Net Exports</i></p> <ul style="list-style-type: none"> • Decrease in foreign demand • Relative price increase of U.S. goods 	<p><i>Net Exports</i></p> <ul style="list-style-type: none"> • Increase in foreign demand • Relative price drop of U.S. goods

Determinants of Aggregate Demand

Key Concepts and Summary

Aggregate demand is the sum of four components: consumption, investment, government spending, and net exports. Consumption will change for a number of reasons, including movements in income, taxes, expectations about future income, and changes in wealth levels. Investment will change in response to its expected profitability, which in turn is shaped by expectations about future economic growth, the creation of new technologies, the price of key inputs, and tax incentives for investment. Investment will also change when interest rates rise or fall. Government spending and taxes are determined by political considerations. Exports and imports change according to relative growth rates and prices between two economies.

Self-Check Questions

Exercise:

Problem:

In the Keynesian framework, which of the following events might cause a recession? Which might cause inflation? Sketch AD/AS diagrams to illustrate your answers.

- a. A large increase in the price of the homes people own.
- b. Rapid growth in the economy of a major trading partner.
- c. The development of a major new technology offers profitable opportunities for business.
- d. The interest rate rises.
- e. The good imported from a major trading partner become much less expensive.

Solution:

- a. An increase in home values will increase consumption spending (due to increased wealth). AD will shift to the right and may cause inflation if it goes beyond potential GDP.
- b. Rapid growth by a major trading partner will increase demand for exports. AD will shift to the right and may cause inflation if it goes beyond potential GDP.
- c. Increased profit opportunities will increase business investment. AD will shift to the right and may cause inflation if it goes beyond potential GDP.
- d. Higher interest rates reduce investment spending. AD will shift to the left and may cause recession if it falls below potential GDP.
- e. Demand for cheaper imports increases, reducing demand for domestic products. AD will shift to the left and may be recessionary.

Exercise:

Problem:

In a Keynesian framework, using an AD/AS diagram, which of the following government policy choices offer a possible solution to recession? Which offer a possible solution to inflation?

- a. A tax increase on consumer income.
- b. A surge in military spending.
- c. A reduction in taxes for businesses that increase investment.
- d. A major increase in what the U.S. government spends on healthcare.

Solution:

- a. A tax increase on consumer income will cause consumption to fall, pushing the AD curve left, and is a possible solution to inflation.
- b. A surge in military spending is an increase in government spending. This will cause the AD curve to shift to the right. If real GDP is less than potential GDP, then this spending would pull the

economy out of a recession. If real GDP is to the right of potential GDP, then the AD curve will shift farther to the right and military spending will be inflationary.

- c. A tax cut focused on business investment will shift AD to the right. If the original macroeconomic equilibrium is below potential GDP, then this policy can help move an economy out of a recession.
- d. Government spending on healthcare will cause the AD curve to shift to the right. If real GDP is less than potential GDP, then this spending would pull the economy out of a recession. If real GDP is to the right of potential GDP, then the AD curve will shift farther to the right and healthcare spending will be inflationary.

Review Questions

Exercise:

Problem:

Name some economic events not related to government policy that could cause aggregate demand to shift.

Exercise:

Problem:

Name some government policies that could cause aggregate demand to shift.

Critical Thinking Questions

Exercise:

Problem:

In its recent report, The Conference Board's *Global Economic Outlook 2015*, updated November 2014 (<http://www.conference-board.org/data/globaloutlook.cfm>), projects China's growth between 2015 and 2019 to be about 5.5%. *International Business Times* (<http://www.ibtimes.com/us-exports-china-have-grown-294-over-past-decade-1338693>) reports that China is the United States' third largest export market, with exports to China growing 294% over the last ten years. Explain what impact China has on the U.S. economy.

Exercise:**Problem:**

What may happen if growth in China continues or contracts?

References

Mahapatra, Lisa. "US Exports To China Have Grown 294% Over The Past Decade." *International Business Times*. Last modified July 09, 2013. <http://www.ibtimes.com/us-exports-china-have-grown-294-over-past-decade-1338693>.

The Conference Board, Inc. "Global Economic Outlook 2014, November 2013." <http://www.conference-board.org/data/globaloutlook.cfm>.

Thomas, G. Scott. "Recession claimed 170,000 small businesses in two years." *The Business Journals*. Last modified July 24, 2012. <http://www.bizjournals.com/bizjournals/on-numbers/scott-thomas/2012/07/recession-claimed-170000-small.html>.

United States Department of Labor: Bureau of Labor Statistics. "Top Picks." <http://data.bls.gov/cgi-bin/surveymost?bls>.

Glossary

disposable income
income after taxes

inflationary gap
equilibrium at a level of output above potential GDP

real GDP
the amount of goods and services actually being sold in a nation

recessionary gap
equilibrium at a level of output below potential GDP

The Building Blocks of Keynesian Analysis

By the end of this section, you will be able to:

- Evaluate the Keynesian view of recessions through an understanding of sticky wages and prices and the importance of aggregate demand
- Explain the coordination argument, menu costs, and macroeconomic externality
- Analyze the impact of the expenditure multiplier

Now that we have a clear understanding of what constitutes aggregate demand, we return to the Keynesian argument using the model of aggregate demand/aggregate supply (AD/AS). (For a similar treatment using Keynes' income-expenditure model, see the appendix on [The Expenditure-Output Model](#).)

Keynesian economics focuses on explaining why recessions and depressions occur and offering a policy prescription for minimizing their effects. The Keynesian view of recession is based on two key building blocks. First, aggregate demand is not always automatically high enough to provide firms with an incentive to hire enough workers to reach full employment. Second, the macroeconomy may adjust only slowly to shifts in aggregate demand because of **sticky wages and prices**, which are wages and prices that do not respond to decreases or increases in demand. We will consider these two claims in turn, and then see how they are represented in the AD/AS model.

The first building block of the Keynesian diagnosis is that recessions occur when the level of household and business sector demand for goods and services is less than what is produced when labor is fully employed. In other words, the intersection of aggregate supply and aggregate demand occurs at a level of output less than the level of GDP consistent with full employment. Suppose the stock market crashes, as occurred in 1929. Or, suppose the housing market collapses, as occurred in 2008. In either case, household wealth will decline, and consumption expenditure will follow. Suppose businesses see that consumer spending is falling. That will reduce expectations of the profitability of investment, so businesses will decrease investment expenditure.

This seemed to be the case during the Great Depression, since the physical capacity of the economy to supply goods did not alter much. No flood or earthquake or other natural disaster ruined factories in 1929 or 1930. No outbreak of disease decimated the ranks of workers. No key input price, like the price of oil, soared on world markets. The U.S. economy in 1933 had just about the same factories, workers, and state of technology as it had had four years earlier in 1929—and yet the economy had shrunk dramatically. This also seems to be what happened in 2008.

As Keynes recognized, the events of the Depression contradicted Say's law that "supply creates its own demand." Although production capacity existed, the markets were not able to sell their products. As a result, real GDP was less than potential GDP.

Wage and Price Stickiness

Keynes also pointed out that although AD fluctuated, prices and wages did not immediately respond as economists often expected. Instead, prices and wages are "sticky," making it difficult to restore the economy to full employment and potential GDP. Keynes emphasized one particular reason why wages were sticky: the **coordination argument**. This argument points out that, even if most people would be willing—at least hypothetically—to see a decline in their own wages in bad economic times as long as everyone else also experienced such a decline, a market-oriented economy has no obvious way to implement a plan of coordinated wage reductions.

[Unemployment](#) proposed a number of reasons why wages might be sticky downward, most of which center on the argument that businesses avoid wage cuts because they may in one way or another depress morale and hurt the productivity of the existing workers.

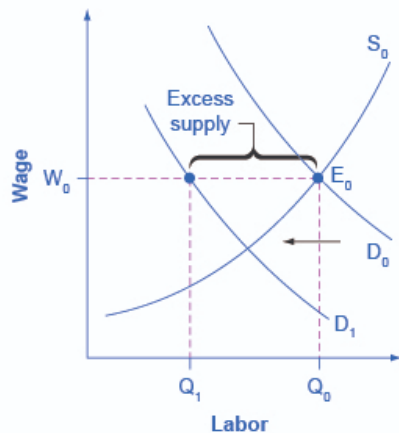
Some modern economists have argued in a Keynesian spirit that, along with wages, other prices may be sticky, too. Many firms do not change their prices every day or even every month. When a firm considers changing prices, it must consider two sets of costs. First, changing prices uses company resources: managers must analyze the competition and market demand and decide what the new prices will be, sales materials must be updated, billing records will change, and product labels and price labels

must be redone. Second, frequent price changes may leave customers confused or angry—especially if they find out that a product now costs more than expected. These costs of changing prices are called **menu costs**—like the costs of printing up a new set of menus with different prices in a restaurant. Prices do respond to forces of supply and demand, but from a macroeconomic perspective, the process of changing all prices throughout the economy takes time.

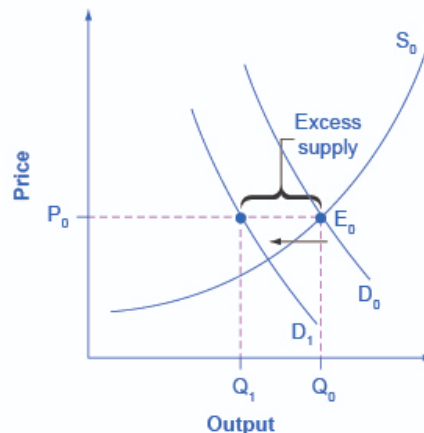
To understand the effect of sticky wages and prices in the economy, consider [\[link\]](#) (a) illustrating the overall labor market, while [\[link\]](#) (b) illustrates a market for a specific good or service. The original equilibrium (E_0) in each market occurs at the intersection of the demand curve (D_0) and supply curve (S_0). When aggregate demand declines, the demand for labor shifts to the left (to D_1) in [\[link\]](#) (a) and the demand for goods shifts to the left (to D_1) in [\[link\]](#) (b). However, because of sticky wages and prices, the wage remains at its original level (W_0) for a period of time and the price remains at its original level (P_0).

As a result, a situation of excess supply—where the quantity supplied exceeds the quantity demanded at the existing wage or price—exists in markets for both labor and goods, and Q_1 is less than Q_0 in both [\[link\]](#) (a) and [\[link\]](#) (b). When many labor markets and many goods markets all across the economy find themselves in this position, the economy is in a recession; that is, firms cannot sell what they wish to produce at the existing market price and do not wish to hire all who are willing to work at the existing market wage. The Clear It Up feature discusses this problem in more detail.

Sticky Prices and Falling Demand in the Labor and Goods Market



(a) Sticky wages in the labor market



(b) Sticky prices in the goods market

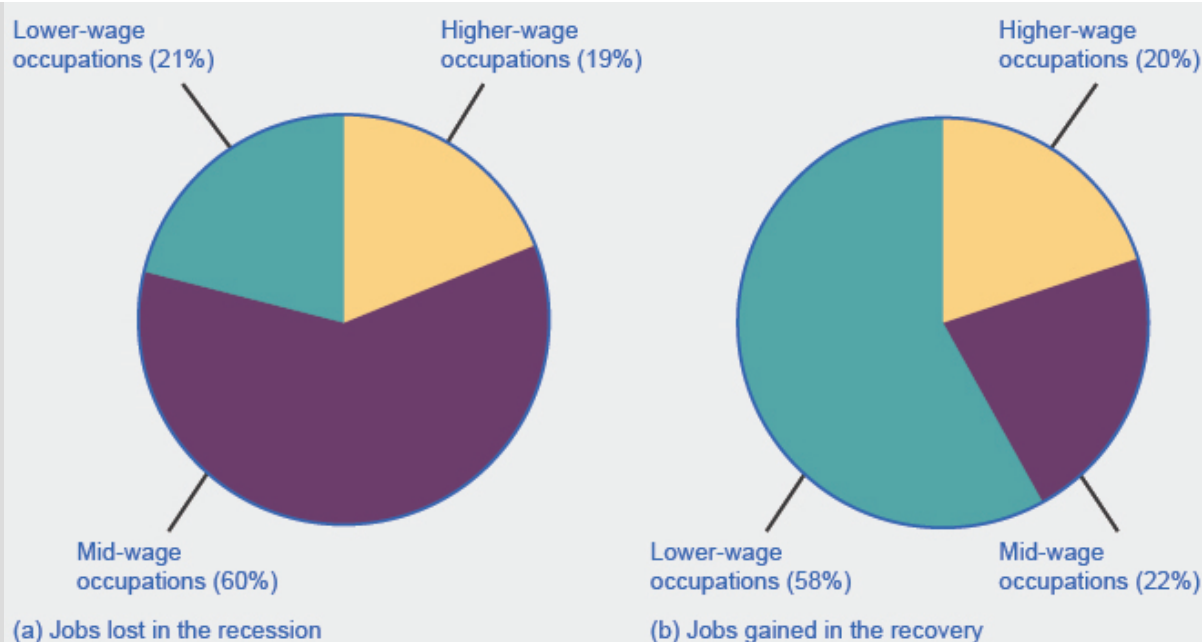
In both (a) and (b), demand shifts left from D_0 to D_1 . However, the wage in (a) and the price in (b) do not immediately decline. In (a), the quantity demanded of labor at the original wage (W_0) is Q_0 , but with the new demand curve for labor (D_1), it will be Q_1 . Similarly, in (b), the quantity demanded of goods at the original price (P_0) is Q_0 , but at the new demand curve (D_1) it will be Q_1 . An excess supply of labor will exist, which is called unemployment. An excess supply of goods will also exist, where the quantity demanded is substantially less than the quantity supplied. Thus, sticky wages and sticky prices, combined with a drop in demand, bring about unemployment and recession.

Note:

Why Is the Pace of Wage Adjustments Slow?

The recovery after the Great Recession in the United States has been slow, with wages stagnant, if not declining. In fact, many low-wage workers at McDonalds, Dominos, and Walmart have threatened to strike for higher wages. Their plight is part of a larger trend in job growth and pay in the post-recession recovery.

Jobs Lost/Gained in the Recession/Recovery



Data in the aftermath of the Great Recession suggests that jobs lost were in mid-wage occupations, while jobs gained were in low-wage occupations.

The National Employment Law Project compiled data from the Bureau of Labor Statistics and found that, during the Great Recession, 60% of job losses were in medium-wage occupations. Most of them were replaced during the recovery period with lower-wage jobs in the service, retail, and food industries. This data is illustrated in [\[link\]](#).

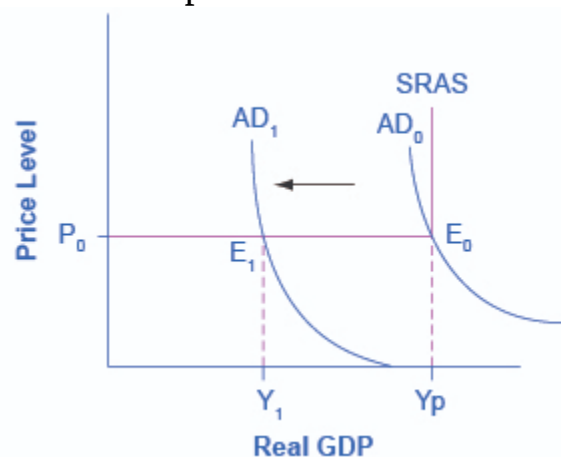
Wages in the service, retail, and food industries are at or near minimum wage and tend to be both downwardly and upwardly “sticky.” Wages are downwardly sticky due to minimum wage laws; they may be upwardly sticky if insufficient competition in low-skilled labor markets enables employers to avoid raising wages that would reduce their profits. At the same time, however, the Consumer Price Index increased 11% between 2007 and 2012, pushing real wages down.

The Two Keynesian Assumptions in the AD/AS Model

These two Keynesian assumptions—the importance of aggregate demand in causing recession and the stickiness of wages and prices—are illustrated by the AD/AS diagram in [\[link\]](#). Note that because of the stickiness of wages and prices, the aggregate supply curve is flatter than either supply curve (labor or specific good). In fact, if wages and prices were so sticky that they did not fall at all, the aggregate supply curve would be completely flat below potential GDP, as shown in [\[link\]](#). This outcome is an important example of a **macroeconomic externality**, where what happens at the macro level is different from and inferior to what happens at the micro level. For example, a firm should respond to a decrease in demand for its product by cutting its price to increase sales. But if all firms experience a decrease in demand for their products, sticky prices in the aggregate prevent aggregate demand from rebounding (which would be shown as a movement along the AD curve in response to a lower price level).

The original equilibrium of this economy occurs where the aggregate demand function (AD_0) intersects with AS. Since this intersection occurs at potential GDP (Y_p), the economy is operating at full employment. When aggregate demand shifts to the left, all the adjustment occurs through decreased real GDP. There is no decrease in the price level. Since the equilibrium occurs at Y_1 , the economy experiences substantial unemployment.

A Keynesian Perspective of Recession



The equilibrium (E_0) illustrates the two key assumptions behind Keynesian economics.

The importance of aggregate demand is

shown because this equilibrium is a recession which has occurred because aggregate demand is at AD_1 instead of AD_0 . The importance of sticky wages and prices is shown because of the assumption of fixed wages and prices, which make the SRAS curve flat below potential GDP. Thus, when AD falls, the intersection E_1 occurs in the flat portion of the SRAS curve where the price level does not change.

The Expenditure Multiplier

A key concept in Keynesian economics is the **expenditure multiplier**. The expenditure multiplier is the idea that not only does spending affect the equilibrium level of GDP, but that spending is powerful. More precisely, it means that a change in spending causes a more than proportionate change in GDP (Y).

Equation:

$$\frac{\Delta Y}{\Delta \text{Spending}} > 1$$

The reason for the expenditure multiplier is that one person's spending becomes another person's income, which leads to additional spending and additional income, and so forth, so that the cumulative impact on GDP is larger than the initial increase in spending. The details of the multiplier process are provided in the appendix on [The Expenditure-Output Model](#), but the concept is important enough to be summarized here. While the multiplier is important for understanding the effectiveness of fiscal policy, it occurs whenever any autonomous increase in spending occurs.

Additionally, the multiplier operates in a negative as well as a positive direction. Thus, when investment spending collapsed during the Great

Depression, it caused a much larger decrease in real GDP. The size of the multiplier is critical and was a key element in recent discussions of the effectiveness of the Obama administration's fiscal stimulus package, officially titled the American Recovery and Reinvestment Act of 2009.

Key Concepts and Summary

Keynesian economics is based on two main ideas: (1) aggregate demand is more likely than aggregate supply to be the primary cause of a short-run economic event like a recession; (2) wages and prices can be sticky, and so, in an economic downturn, unemployment can result. The latter is an example of a macroeconomic externality. While surpluses cause prices to fall at the micro level, they do not necessarily at the macro level; instead the adjustment to a decrease in demand occurs only through decreased quantities. One reason why prices may be sticky is menu costs, the costs of changing prices. These include internal costs a business faces in changing prices in terms of labeling, recordkeeping, and accounting, and also the costs of communicating the price change to (possibly unhappy) customers. Keynesians also believe in the existence of the expenditure multiplier—the notion that a change in autonomous expenditure causes a more than proportionate change in GDP.

Self-Check Questions

Exercise:

Problem:

Use the AD/AS model to explain how an inflationary gap occurs, beginning from the initial equilibrium in [\[link\]](#).

Solution:

An inflationary gap is the result of an increase in aggregate demand when the economy is at potential output. Since the AS curve is vertical at potential GDP, any increase in AD will lead to a higher price level

(i.e. inflation) but no higher real GDP. This is easy to see if you draw AD_1 to the right of AD_0 .

Exercise:

Problem:

Suppose the U.S. Congress cuts federal government spending in order to balance the Federal budget. Use the AD/AS model to analyze the likely impact on output and employment. *Hint:* revisit [\[link\]](#).

Solution:

A decrease in government spending will shift AD to the left.

Review Questions

Exercise:

Problem:

From a Keynesian point of view, which is more likely to cause a recession: aggregate demand or aggregate supply, and why?

Exercise:

Problem:

Why do sticky wages and prices increase the impact of an economic downturn on unemployment and recession?

Exercise:

Problem: Explain what economists mean by “menu costs.”

Critical Thinking Questions

Exercise:

Problem:

Does it make sense that wages would be sticky downwards but not upwards? Why or why not?

Exercise:**Problem:**

Suppose the economy is operating at potential GDP when it experiences an increase in export demand. How might the economy increase production of exports to meet this demand, given that the economy is already at full employment?

References

Harford, Tim. “What Price Supply and Demand?”

http://timharford.com/2014/01/what-price-supply-and-demand/?utm_source=dlvr.it&utm_medium=twitter.

National Employment Law Project. “Job Creation and Economic Recovery.”

http://www.nelp.org/index.php/content/content_issues/category/job_creation_and_economic_recovery/.

Glossary

coordination argument

downward wage and price flexibility requires perfect information about the level of lower compensation acceptable to other laborers and market participants

expenditure multiplier

Keynesian concept that asserts that a change in autonomous spending causes a more than proportionate change in real GDP

macroeconomic externality

occurs when what happens at the macro level is different from and inferior to what happens at the micro level; an example would be where upward sloping supply curves for firms become a flat aggregate supply curve, illustrating that the price level cannot fall to stimulate aggregate demand

menu costs

costs firms face in changing prices

sticky wages and prices

a situation where wages and prices do not fall in response to a decrease in demand, or do not rise in response to an increase in demand

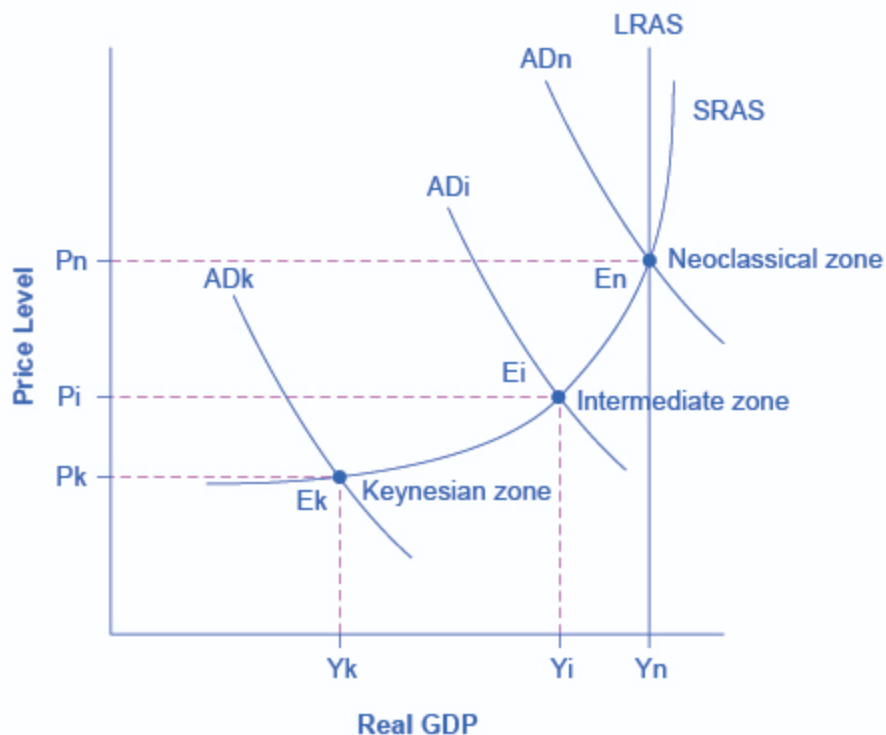
The Phillips Curve

By the end of this section, you will be able to:

- Explain the Phillips curve, noting its impact on the theories of Keynesian economics
- Graph a Phillips curve
- Identify factors that cause the instability of the Phillips curve
- Analyze the Keynesian policy for reducing unemployment and inflation

The simplified AD/AS model that we have used so far is fully consistent with Keynes's original model. More recent research, though, has indicated that in the real world, an aggregate supply curve is more curved than the right angle used in this chapter. Rather, the real-world AS curve is very flat at levels of output far below potential ("the Keynesian zone"), very steep at levels of output above potential ("the neoclassical zone") and curved in between ("the intermediate zone"). This is illustrated in [\[link\]](#). The typical aggregate supply curve leads to the concept of the Phillips curve.

Keynes, Neoclassical, and Intermediate Zones in the Aggregate Supply Curve



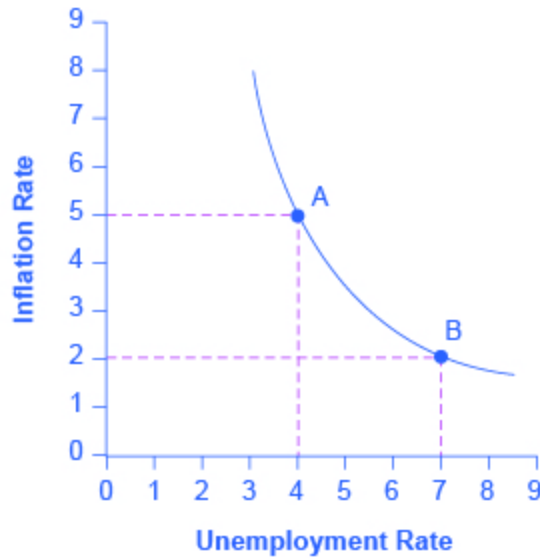
Near the equilibrium E_k , in the Keynesian zone at the far left of the SRAS curve, small shifts in AD, either to the right or the left, will affect the output level Y_k , but will not much affect the price level.

In the Keynesian zone, AD largely determines the quantity of output. Near the equilibrium E_n , in the neoclassical zone, at the far right of the SRAS curve, small shifts in AD, either to the right or the left, will have relatively little effect on the output level Y_n , but instead will have a greater effect on the price level. In the neoclassical zone, the near-vertical SRAS curve close to the level of potential GDP (as represented by the LRAS line) largely determines the quantity of output. In the intermediate zone around equilibrium E_i , movement in AD to the right will increase both the output level and the price level, while a movement in AD to the left would decrease both the output level and the price level.

The Discovery of the Phillips Curve

In the 1950s, A.W. Phillips, an economist at the London School of Economics, was studying the Keynesian analytical framework. The Keynesian theory implied that during a recession inflationary pressures are low, but when the level of output is at or even pushing beyond potential GDP, the economy is at greater risk for inflation. Phillips analyzed 60 years of British data and did find that tradeoff between unemployment and inflation, which became known as a **Phillips curve**. [\[link\]](#) shows a theoretical Phillips curve, and the following Work It Out feature shows how the pattern appears for the United States.

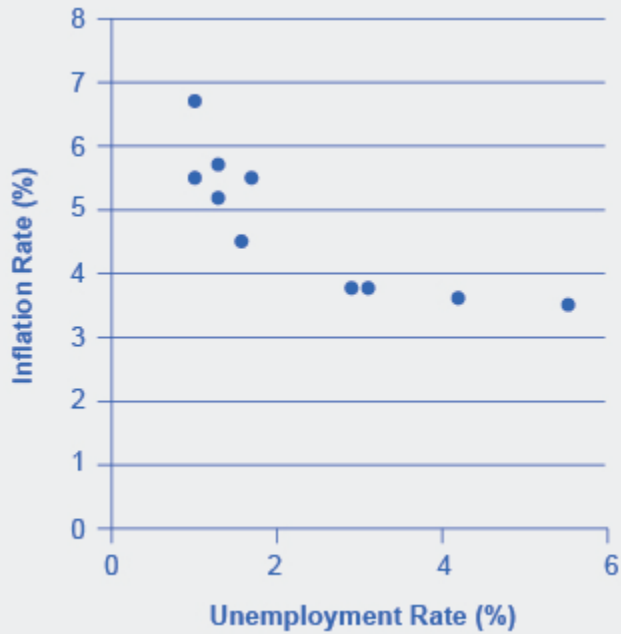
A Keynesian Phillips Curve Tradeoff between Unemployment and Inflation



A Phillips curve illustrates a tradeoff between the unemployment rate and the inflation rate; if one is higher, the other must be lower. For example, point A illustrates an inflation rate of 5% and an unemployment rate of 4%. If the government attempts to reduce inflation to 2%, then it will experience a rise in unemployment to 7%, as shown at point B.

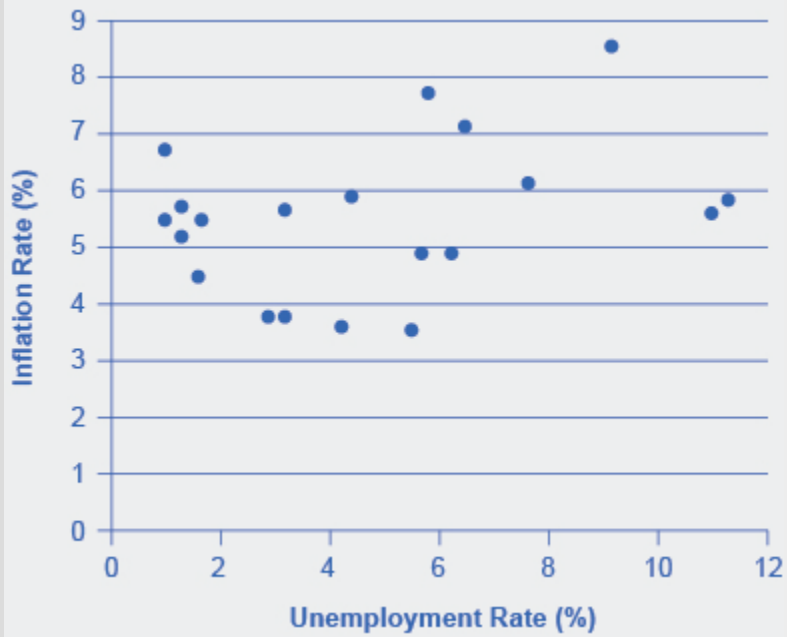
Note:

The Phillips Curve for the United States
The Phillips Curve from 1960–1969



This chart shows the negative relationship between unemployment and inflation.

U.S. Phillips Curve, 1960–1979



The tradeoff between unemployment and

inflation appeared to break down during the 1970s as the Phillips Curve shifted out to the right.

Over this longer period of time, the Phillips curve appears to have shifted out. There is no tradeoff any more.

The Instability of the Phillips Curve

During the 1960s, the Phillips curve was seen as a policy menu. A nation could choose low inflation and high unemployment, or high inflation and low unemployment, or anywhere in between. Fiscal and monetary policy could be used to move up or down the Phillips curve as desired. Then a curious thing happened. When policymakers tried to exploit the tradeoff between inflation and unemployment, the result was an increase in both inflation and unemployment. What had happened? The Phillips curve shifted.

The U.S. economy experienced this pattern in the deep recession from 1973 to 1975, and again in back-to-back recessions from 1980 to 1982. Many nations around the world saw similar increases in unemployment and inflation. This pattern became known as stagflation. (Recall from [The Aggregate Demand/Aggregate Supply Model](#) that stagflation is an unhealthy combination of high unemployment and high inflation.) Perhaps most important, stagflation was a phenomenon that could not be explained by traditional Keynesian economics.

Economists have concluded that two factors cause the Phillips curve to shift. The first is supply shocks, like the Oil Crisis of the mid-1970s, which first brought stagflation into our vocabulary. The second is changes in people's expectations about inflation. In other words, there may be a tradeoff between inflation and unemployment when people expect no inflation, but when they realize inflation is occurring, the tradeoff disappears. Both factors (supply shocks and changes in inflationary

expectations) cause the aggregate supply curve, and thus the Phillips curve, to shift.

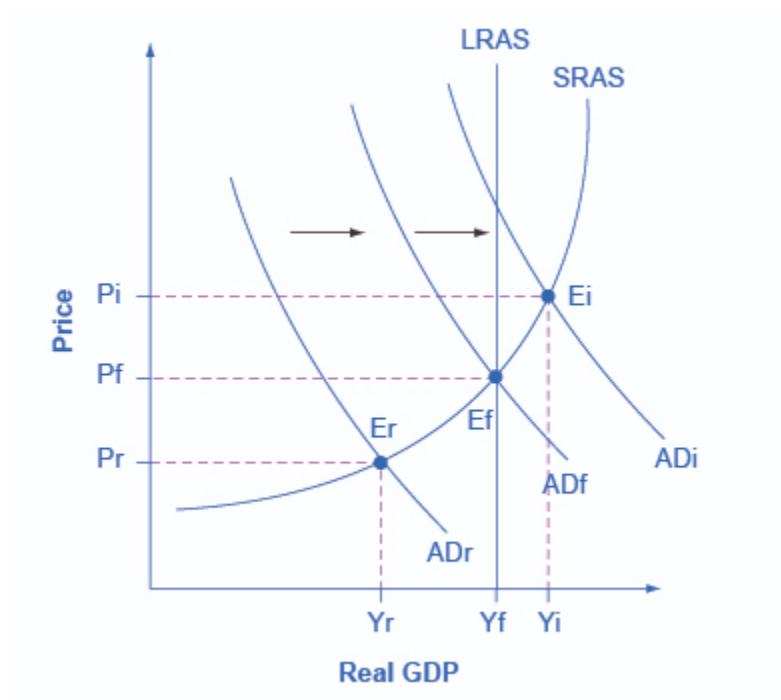
In short, a downward-sloping Phillips curve should be interpreted as valid for short-run periods of several years, but over longer periods, when aggregate supply shifts, the downward-sloping Phillips curve can shift so that unemployment and inflation are both higher (as in the 1970s and early 1980s) or both lower (as in the early 1990s or first decade of the 2000s).

Keynesian Policy for Fighting Unemployment and Inflation

Keynesian macroeconomics argues that the solution to a recession is **expansionary fiscal policy**, such as tax cuts to stimulate consumption and investment, or direct increases in government spending that would shift the aggregate demand curve to the right. For example, if aggregate demand was originally at AD_r in [\[link\]](#), so that the economy was in recession, the appropriate policy would be for government to shift aggregate demand to the right from AD_r to AD_f, where the economy would be at potential GDP and full employment.

Keynes noted that while it would be nice if the government could spend additional money on housing, roads, and other amenities, he also argued that if the government could not agree on how to spend money in practical ways, then it could spend in impractical ways. For example, Keynes suggested building monuments, like a modern equivalent of the Egyptian pyramids. He proposed that the government could bury money underground, and let mining companies get started to dig the money up again. These suggestions were slightly tongue-in-cheek, but their purpose was to emphasize that a Great Depression is no time to quibble over the specifics of government spending programs and tax cuts when the goal should be to pump up aggregate demand by enough to lift the economy to potential GDP.

Fighting Recession and Inflation with Keynesian Policy



If an economy is in recession, with an equilibrium at E_r , then the Keynesian response would be to enact a policy to shift aggregate demand to the right from AD_r toward AD_f . If an economy is experiencing inflationary pressures with an equilibrium at E_i , then the Keynesian response would be to enact a policy response to shift aggregate demand to the left, from AD_i toward AD_f .

The other side of Keynesian policy occurs when the economy is operating above potential GDP. In this situation, unemployment is low, but inflationary rises in the price level are a concern. The Keynesian response would be **contractionary fiscal policy**, using tax increases or government spending cuts to shift AD to the left. The result would be downward pressure on the price level, but very little reduction in output or very little rise in unemployment. If aggregate demand was originally at AD_i in [\[link\]](#), so that the economy was experiencing inflationary rises in the price level,

the appropriate policy would be for government to shift aggregate demand to the left, from AD_i toward AD_f, which reduces the pressure for a higher price level while the economy remains at full employment.

In the Keynesian economic model, too little aggregate demand brings unemployment and too much brings inflation. Thus, you can think of Keynesian economics as pursuing a “Goldilocks” level of aggregate demand: not too much, not too little, but looking for what is just right.

Key Concepts and Summary

A Phillips curve shows the tradeoff between unemployment and inflation in an economy. From a Keynesian viewpoint, the Phillips curve should slope down so that higher unemployment means lower inflation, and vice versa. However, a downward-sloping Phillips curve is a short-term relationship that may shift after a few years.

Keynesian macroeconomics argues that the solution to a recession is expansionary fiscal policy, such as tax cuts to stimulate consumption and investment, or direct increases in government spending that would shift the aggregate demand curve to the right. The other side of Keynesian policy occurs when the economy is operating above potential GDP. In this situation, unemployment is low, but inflationary rises in the price level are a concern. The Keynesian response would be contractionary fiscal policy, using tax increases or government spending cuts to shift AD to the left.

Self-Check Question

Exercise:

Problem:

How would a decrease in energy prices affect the Phillips curve?

Solution:

A decrease in energy prices, a positive supply shock, would cause the AS curve to shift out to the right, yielding more real GDP at a lower price level. This would shift the Phillips curve down toward the origin, meaning the economy would experience lower unemployment and a lower rate of inflation.

Review Questions

Exercise:

Problem: What tradeoff is shown by a Phillips curve?

Exercise:

Problem:

Would you expect to see long-run data trace out a stable downward-sloping Phillips curve?

Exercise:

Problem:

What is the Keynesian prescription for recession? For inflation?

Critical Thinking Questions

Exercise:

Problem:

Do you think the Phillips curve is a useful tool for analyzing the economy today? Why or why not?

References

Hoover, Kevin. "Phillips Curve." *The Concise Encyclopedia of Economics*.
<http://www.econlib.org/library/Enc/PhillipsCurve.html>.

U.S. Government Printing Office. "Economic Report of the President."
<http://1.usa.gov/1c3psdL>.

Glossary

contractionary fiscal policy

tax increases or cuts in government spending designed to decrease aggregate demand and reduce inflationary pressures

expansionary fiscal policy

tax cuts or increases in government spending designed to stimulate aggregate demand and move the economy out of recession

Phillips curve

the tradeoff between unemployment and inflation

The Keynesian Perspective on Market Forces

By the end of this section, you will be able to:

- Explain the Keynesian perspective on market forces
- Analyze the role of government policy in economic management

Ever since the birth of Keynesian economics in the 1930s, controversy has simmered over the extent to which government should play an active role in managing the economy. In the aftermath of the human devastation and misery of the Great Depression, many people—including many economists—became more aware of vulnerabilities within the market-oriented economic system. Some supporters of Keynesian economics advocated a high degree of government planning in all parts of the economy.

However, Keynes himself was careful to separate the issue of aggregate demand from the issue of how well individual markets worked. He argued that individual markets for goods and services were appropriate and useful, but that sometimes that level of aggregate demand was just too low. When 10 million people are willing and able to work, but one million of them are unemployed, he argued, individual markets may be doing a perfectly good job of allocating the efforts of the nine million workers—the problem is that insufficient aggregate demand exists to support jobs for all 10 million. Thus, he believed that, while government should ensure that overall level of aggregate demand is sufficient for an economy to reach full employment, this task did not imply that the government should attempt to set prices and wages throughout the economy, nor to take over and manage large corporations or entire industries directly.

Even if one accepts the Keynesian economic theory, a number of practical questions remain. In the real world, can government economists identify potential GDP accurately? Is a desired increase in aggregate demand better accomplished by a tax cut or by an increase in government spending? Given the inevitable delays and uncertainties as policies are enacted into law, is it reasonable to expect that the government can implement Keynesian economics? Can fixing a recession really be just as simple as pumping up aggregate demand? [Government Budgets and Fiscal Policy](#) will probe these issues. The Keynesian approach, with its focus on aggregate demand and sticky prices, has proved useful in understanding how the economy

fluctuates in the short run and why recessions and cyclical unemployment occur. In [The Neoclassical Perspective](#), we will consider some of the shortcomings of the Keynesian approach and why it is not especially well-suited for long-run macroeconomic analysis.

Note:

The Great Recession

The lessons learned during the Great Depression of the 1930s and the aggregate expenditure model proposed by John Maynard Keynes gave the modern economists and policymakers of today the tools to effectively navigate the treacherous economy in the latter half of the 2000s. In “How the Great Recession Was Brought to an End,” Alan S. Blinder and Mark Zandi wrote that the actions taken by today’s policymakers stand in sharp contrast to those of the early years of the Great Depression. Today’s economists and policymakers were not content to let the markets recover from recession without taking proactive measures to support consumption and investment. The Federal Reserve actively lowered short-term interest rates and developed innovative ways to pump money into the economy so that credit and investment would not dry up. Both Presidents Bush and Obama and Congress implemented a variety of programs ranging from tax rebates to “Cash for Clunkers” to the Troubled Asset Relief Program to stimulate and stabilize household consumption and encourage investment. Although these policies came under harsh criticism from the public and many politicians, they lessened the impact of the economic downturn and may have saved the country from a second Great Depression.

Key Concepts and Summary

The Keynesian prescription for stabilizing the economy implies government intervention at the macroeconomic level—increasing aggregate demand when private demand falls and decreasing aggregate demand when private demand rises. This does not imply that the government should be passing laws or regulations that set prices and quantities in microeconomic markets.

Self-Check Questions

Exercise:

Problem:

Does Keynesian economics require government to set controls on prices, wages, or interest rates?

Solution:

Keynesian economics does not require microeconomic price controls of any sort. It is true that many Keynesian economic prescriptions were for the government to influence the total amount of aggregate demand in the economy, often through government spending and tax cuts.

Exercise:

Problem:

List three practical problems with the Keynesian perspective.

Solution:

The three problems center on government's ability to estimate potential GDP, decide whether to influence aggregate demand through tax changes or changes in government spending, and the lag time that occurs as Congress and the President attempt to pass legislation.

Review Questions

Exercise:

Problem:

How did the Keynesian perspective address the economic market failure of the Great Depression?

Critical Thinking Questions

Exercise:

Problem:

Return to the table from the *Economic Report of the President* in the earlier [Work It Out](#) feature titled “The Phillips Curve for the United States.” How would you expect government spending to have changed over the last six years?

Exercise:

Problem:

Explain what types of policies the federal government may have implemented to restore aggregate demand and the potential obstacles policymakers may have encountered.

References

Blinder, Alan S., and Mark Zandi. “How the Great Recession Was Brought to an End.” Last modified July 27, 2010.
<http://www.princeton.edu/~blinder/End-of-Great-Recession.pdf>.

Introduction to the Neoclassical Perspective

class="introduction"

Impact of the Great Recession

The impact of the Great Recession can be seen in many areas of the economy that impact our daily lives. One of the most visible signs can be seen in the housing market where many homes and other buildings are abandoned, including ones that midway through construction.

(Credit: modification of work by A McLin/Flickr r Creative Commons)

**Note:****Navigating Uncharted Waters**

The Great Recession ended in June 2009 after 18 months, according to the National Bureau of Economic Research (NBER). The NBER examines a variety of measures of economic activity to gauge the overall health of the economy. These measures include real income, wholesale and retail sales, employment, and industrial production. In the years since the official end of this historic economic downturn, it has become clear that the Great Recession was two-pronged, hitting the U.S. economy with the collapse of the housing market and the failure of the financial system's credit institutions, further contaminating global economies. While the stock market rapidly lost trillions of dollars of value, consumer spending dried up, and companies began cutting jobs, economic policymakers were struggling with how to best combat and prevent a national, and even global economic collapse. In the end, policymakers used a number of controversial monetary and fiscal policies to support the housing market and domestic industries as well as to stabilize the financial sector. Some of these initiatives included:

- Federal Reserve Bank purchase of both traditional and nontraditional assets off banks' balance sheets. By doing this, the Fed injected money into the banking system and increased the amounts of funds available to lend to the business sector and consumers. This also dropped short-term interest rates to as low as zero percent and had the effect of devaluing U.S. dollars in the global market and boosting exports.
- The Congress and the President also passed several pieces of legislation that would stabilize the financial market. The Troubled Asset Relief Program (TARP), passed in late 2008, allowed the government to inject cash into troubled banks and other financial institutions and help support General Motors and Chrysler as they faced bankruptcy and threatened job losses throughout their supply chain. The American Recovery and Reinvestment Act in early 2009 provided tax rebates to low- and middle-income households to encourage consumer spending.

Four years after the end of the Great Recession, the economy has yet to return to its pre-recession levels of productivity and growth. Annual productivity increased only 1.9% between 2009 and 2012 compared to its 2.7% annual growth rate between 2000 and 2007, unemployment remains above the natural rate, and real GDP continues to lag behind potential growth. The actions taken to stabilize the economy are still under scrutiny and debate about their effectiveness continues. In this chapter, we will discuss the neoclassical perspective on economics and compare it to the Keynesian perspective. At the end of the chapter, we will use the neoclassical perspective to analyze the actions taken in the Great Recession.

Note:

Introduction to the Neoclassical Perspective

In this chapter, you will learn about:

- The Building Blocks of Neoclassical Analysis
- The Policy Implications of the Neoclassical Perspective

- Balancing Keynesian and Neoclassical Models

In Chicago, Illinois, the highest recorded temperature was 105° in July 1995, while the lowest recorded temperature was 27° below zero in January 1958. Understanding why these extreme weather patterns occurred would be interesting. However, if you wanted to understand the typical weather pattern in Chicago, instead of focusing on one-time extremes, you would need to look at the entire pattern of data over time.

A similar lesson applies to the study of macroeconomics. It is interesting to study extreme situations, like the Great Depression of the 1930s or what many have called the Great Recession of 2008–2009. If you want to understand the whole picture, however, you need to look at the long term. Consider the unemployment rate. The unemployment rate has fluctuated from as low as 3.5% in 1969 to as high as 9.7% in 1982 and 9.6% in 2009. Even as the U.S. unemployment rate rose during recessions and declined during expansions, it kept returning to the general neighborhood of 5.0–5.5%. When the nonpartisan Congressional Budget Office carried out its long-range economic forecasts in 2010, it assumed that from 2015 to 2020, after the recession has passed, the unemployment rate would be 5.0%. From a long-run perspective, the economy seems to keep adjusting back to this rate of unemployment.

As the name “neoclassical” implies, this perspective of how the macroeconomy works is a “new” view of the “old” classical model of the economy. The classical view, the predominant economic philosophy until the Great Depression, was that short-term fluctuations in economic activity would rather quickly, with flexible prices, adjust back to full employment. This view of the economy implied a vertical aggregate supply curve at full employment GDP, and prescribed a “hands off” policy approach. For example, if the economy were to slip into recession (a leftward shift of the aggregate demand curve), it would temporarily exhibit a surplus of goods. This surplus would be eliminated with falling prices, and the economy would return to full employment level of GDP; no active fiscal or monetary policy was needed. In fact, the classical view was that expansionary fiscal

or monetary policy would only cause inflation, rather than increase GDP. The deep and lasting impact of the Great Depression changed this thinking and Keynesian economics, which prescribed active fiscal policy to alleviate weak aggregate demand, became the more mainstream perspective.

The Building Blocks of Neoclassical Analysis

By the end of this section, you will be able to:

- Explain the importance of potential GDP in the long run
- Analyze the role of flexible prices
- Interpret a neoclassical model of aggregate demand and aggregate supply
- Evaluate different ways for measuring the speed of macroeconomic adjustment

The **neoclassical perspective** on macroeconomics holds that, in the long run, the economy will fluctuate around its potential GDP and its natural rate of unemployment. This chapter begins with two building blocks of neoclassical economics: (1) the size of the economy is determined by potential GDP, and (2) wages and prices will adjust in a flexible manner so that the economy will adjust back to its potential GDP level of output. The key policy implication is this: Should the government focus more on long-term growth and on controlling inflation than on worrying about recession or cyclical unemployment? This focus on long-run growth rather than the short-run fluctuations in the business cycle means that neoclassical economics is more useful for long-run macroeconomic analysis and Keynesian economics is more useful for analyzing the macroeconomic short run. Let's consider the two neoclassical building blocks in turn, and how they can be embodied in the aggregate demand/aggregate supply model.

The Importance of Potential GDP in the Long Run

Over the long run, the level of potential GDP determines the size of real GDP. When economists refer to “potential GDP” they are referring to that level of output that can be achieved when all resources (land, labor, capital, and entrepreneurial ability) are fully employed. While the unemployment rate in labor markets will never be zero, full employment in the labor market refers to zero cyclical unemployment. There will still be some level of unemployment due to frictional or structural unemployment, but when the economy is operating with zero cyclical unemployment, the economy is said to be at the natural rate of unemployment or at full employment.

Actual or real GDP is benchmarked against the potential GDP to determine how well the economy is performing. Growth in GDP can be explained by increases and investment in physical capital and human capital per person as well as advances in technology. **Physical capital per person** refers to the amount and kind of machinery and equipment available to help people get work done. Compare, for example, your productivity in typing a term paper on a typewriter to working on your laptop with word processing software. Clearly, you will be able to be more productive using word processing software. The technology and level of capital of your laptop and software has increased your productivity. More broadly, the development of GPS technology and Universal Product Codes (those barcodes on every product we buy) has made it much easier for firms to track shipments, tabulate inventories, and sell and distribute products. These two technological innovations, and many others, have increased a nation's ability to produce goods and services for a given population. Likewise, increasing human capital involves increasing levels of knowledge, education, and skill sets per person through vocational or higher education. Physical and human capital improvements with technological advances will increase overall productivity and, thus, GDP.

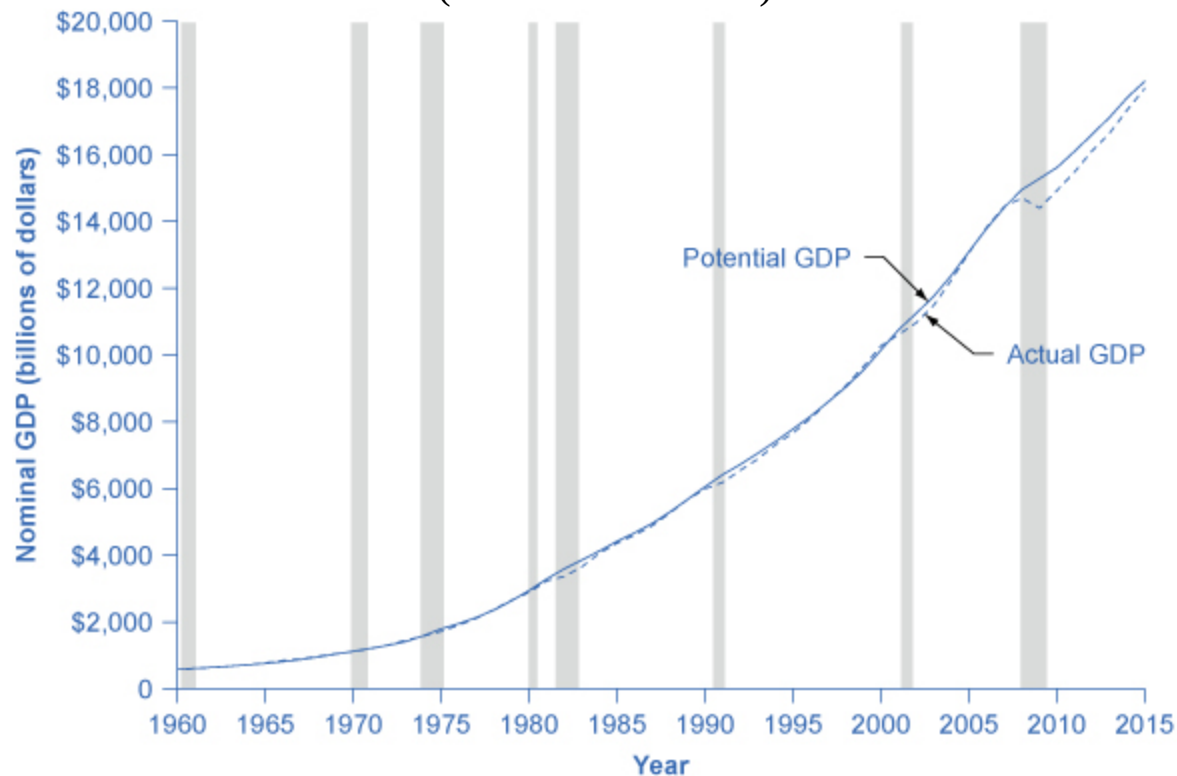
To see how these improvements have increased productivity and output at the national level, we should examine evidence from the United States. The United States experienced significant growth in the twentieth century due to phenomenal changes in infrastructure, equipment, and technological improvements in physical capital and human capital. The population more than tripled in the twentieth century, from 76 million in 1900 to over 300 million in 2012. The human capital of modern workers is far higher today because the education and skills of workers have risen dramatically. In 1900, only about one-eighth of the U.S. population had completed high school and just one person in 40 had completed a four-year college degree. By 2010, more than 87% of Americans had a high school degree and over 29% had a four-year college degree as well. In 2014, 40% of working-age Americans had a four-year college degree. The average amount of physical capital per worker has grown dramatically. The technology available to modern workers is extraordinarily better than a century ago: cars, airplanes, electrical machinery, smartphones, computers, chemical and biological advances, materials science, health care—the list of technological advances

could run on and on. More workers, higher skill levels, larger amounts of physical capital per worker, and amazingly better technology, and potential GDP for the U.S. economy has clearly increased a great deal since 1900.

This growth has fallen below its potential GDP and, at times, has exceeded its potential. For example from 2008 to 2009, the U.S. economy tumbled into recession and remains below its potential. At other times, like in the late 1990s, the economy ran at potential GDP—or even slightly ahead.

[\[link\]](#) shows the actual data for the increase in nominal GDP since 1960. The slightly smoother line shows the potential GDP since 1960 as estimated by the nonpartisan Congressional Budget Office. Most economic recessions and upswings are times when the economy is 1–3% below or above potential GDP in a given year. Clearly, short-run fluctuations around potential GDP do exist, but over the long run, the upward trend of potential GDP determines the size of the economy.

Potential and Actual GDP (in Nominal Dollars)

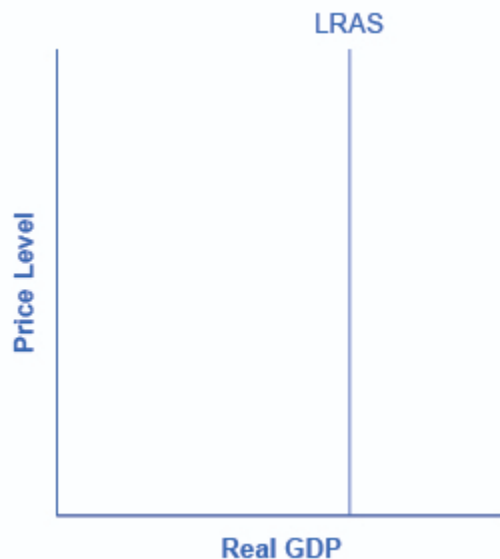


Actual GDP falls below potential GDP during and after recessions (gray shaded areas indicate US recessions). In other cases, actual GDP can be above potential GDP for a time, as in the late 1990s.

(Source: Federal Reserve Economic Data (FRED)
<https://fred.stlouisfed.org/graph/?g=ef0J>)

In the aggregate demand/aggregate supply model, potential GDP is shown as a vertical line. Neoclassical economists who focus on potential GDP as the primary determinant of real GDP argue that the long-run aggregate supply curve is located at potential GDP—that is, the long-run aggregate supply curve is a vertical line drawn at the level of potential GDP, as shown in [\[link\]](#). A vertical LRAS curve means that the level of aggregate supply (or potential GDP) will determine the real GDP of the economy, regardless of the level of aggregate demand. Over time, increases in the quantity and quality of physical capital, increases in human capital, and technological advancements shift potential GDP and the vertical LRAS curve gradually to the right. This gradual increase in an economy's potential GDP is often described as a nation's long-term economic growth.

A Vertical AS Curve



In the neoclassical model, the aggregate supply curve is drawn as a vertical line at the level of potential GDP. If AS is vertical, then it determines the level of real output, no matter where the aggregate demand

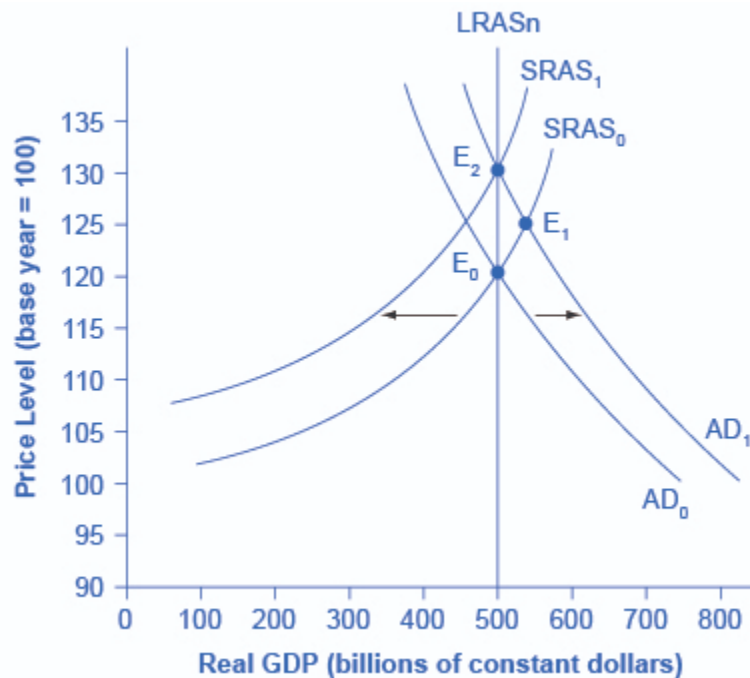
curve is drawn. Over time, the LRAS curve shifts to the right as productivity increases and potential GDP expands.

The Role of Flexible Prices

How does the macroeconomy adjust back to its level of potential GDP in the long run? What if aggregate demand increases or decreases? The neoclassical view of how the macroeconomy adjusts is based on the insight that even if wages and prices are “sticky”, or slow to change, in the short run, they are flexible over time. To understand this better, let's follow the connections from the short-run to the long-run macroeconomic equilibrium.

The aggregate demand and aggregate supply diagram shown in [\[link\]](#) shows two aggregate supply curves. The original upward sloping aggregate supply curve ($SRAS_0$) is a short-run or Keynesian AS curve. The vertical aggregate supply curve ($LRAS_n$) is the long-run or neoclassical AS curve, which is located at potential GDP. The original aggregate demand curve, labeled AD_0 , is drawn so that the original equilibrium occurs at point E_0 , at which point the economy is producing at its potential GDP.

The Rebound to Potential GDP after AD Increases



The original equilibrium (E_0), at an output level of 500 and a price level of 120, happens at the intersection of the aggregate demand curve (AD_0) and the short-run aggregate supply curve ($SRAS_0$). The output at E_0 is equal to potential GDP. Aggregate demand shifts right from AD_0 to AD_1 . The new equilibrium is E_1 , with a higher output level of 550 and an increase in the price level to 125. With unemployment rates unsustainably low, wages are bid up by eager employers, which shifts short-run aggregate supply to the left, from $SRAS_0$ to $SRAS_1$. The new equilibrium (E_2) is at the same original level of output, 500, but at a higher price level of 130. Thus, the long-run aggregate supply curve ($LRAS_n$), which is vertical at the level of potential GDP, determines the level of real GDP in this economy in the long run.

Now, imagine that some economic event boosts aggregate demand: perhaps a surge of export sales or a rise in business confidence that leads to more investment, perhaps a policy decision like higher government spending, or perhaps a tax cut that leads to additional aggregate demand. The short-run Keynesian analysis is that the rise in aggregate demand will shift the aggregate demand curve out to the right, from AD_0 to AD_1 , leading to a new equilibrium at point E_1 with higher output, lower unemployment, and pressure for an inflationary rise in the price level.

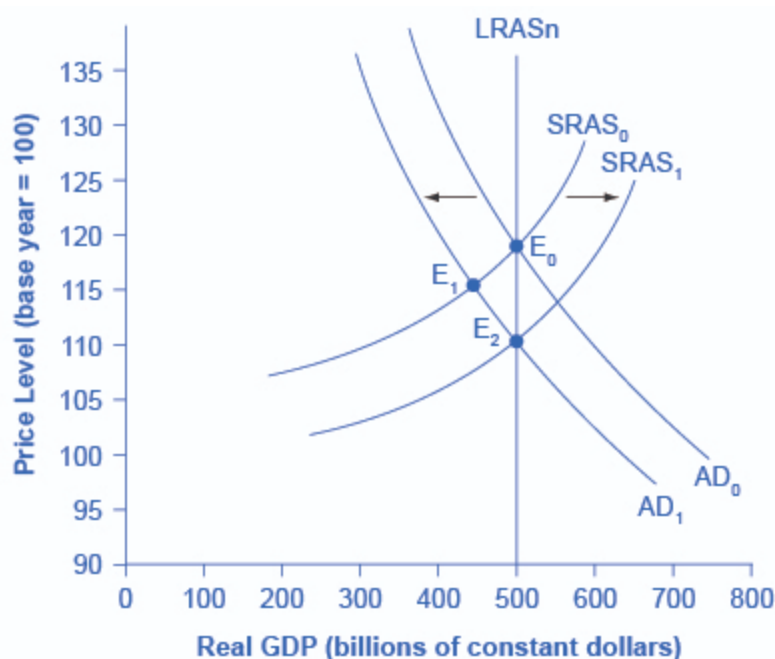
In the long-run neoclassical analysis, however, the chain of economic events is just beginning. As economic output rises above potential GDP, the level of unemployment falls. The economy is now above full employment and there is a shortage of labor. Eager employers are trying to bid workers away from other companies and to encourage their current workers to exert more effort and to put in longer hours. This high demand for labor will drive up wages. Most workers have their salaries reviewed only once or twice a year, and so it will take time before the higher wages filter through the economy. As wages do rise, it will mean a leftward shift in the short-run Keynesian aggregate supply curve back to $SRAS_1$, because the price of a major input to production has increased. The economy moves to a new equilibrium (E_2). The new equilibrium has the same level of real GDP as did the original equilibrium (E_0), but there has been an inflationary increase in the price level.

This description of the short-run shift from E_0 to E_1 and the long-run shift from E_1 to E_2 is a step-by-step way of making a simple point: the economy cannot sustain production above its potential GDP in the long run. An economy may produce above its level of potential GDP in the short run, under pressure from a surge in aggregate demand. Over the long run, however, that surge in aggregate demand ends up as an increase in the price level, not as a rise in output.

The rebound of the economy back to potential GDP also works in response to a shift to the left in aggregate demand. [\[link\]](#) again starts with two aggregate supply curves, with $SRAS_0$ showing the original upward sloping

short-run Keynesian AS curve and LRASn showing the vertical long-run neoclassical aggregate supply curve. A decrease in aggregate demand—for example, because of a decline in consumer confidence that leads to less consumption and more saving—causes the original aggregate demand curve AD_0 to shift back to AD_1 . The shift from the original equilibrium (E_0) to the new equilibrium (E_1) results in a decline in output. The economy is now below full employment and there is a surplus of labor. As output falls below potential GDP, unemployment rises. While a lower price level (i.e., deflation) is rare in the United States, it does happen from time to time during very weak periods of economic activity. For practical purposes, we might consider a lower price level in the AD–AS model as indicative of disinflation, which is a decline in the rate of inflation. Thus, the long-run aggregate supply curve LRASn, which is vertical at the level of potential GDP, ultimately determines the real GDP of this economy.

A Rebound Back to Potential GDP from a Shift to the Left in Aggregate Demand



The original equilibrium (E_0), at an output level of 500 and a price level of 120, happens at the intersection of the aggregate demand curve (AD_0) and the short-run aggregate supply curve ($SRAS_0$). The output at E_0 is equal to potential GDP.

Aggregate demand shifts left, from AD_0 to AD_1 . The new equilibrium is at E_1 , with a lower output level of 450 and downward pressure on the price level of 115. With high unemployment rates, wages are held down. Lower wages are an economy-wide decrease in the price of a key input, which shifts short-run aggregate supply to the right, from $SRAS_0$ to $SRAS_1$. The new equilibrium (E_2) is at the same original level of output, 500, but at a lower price level of 110.

Again, from the neoclassical perspective, this short-run scenario is only the beginning of the chain of events. The higher level of unemployment means more workers looking for jobs. As a result, employers can hold down on pay increases—or perhaps even replace some of their higher-paid workers with unemployed people willing to accept a lower wage. As wages stagnate or fall, this decline in the price of a key input means that the short-run Keynesian aggregate supply curve shifts to the right from its original ($SRAS_0$ to $SRAS_1$). The overall impact in the long run, as the macroeconomic equilibrium shifts from E_0 to E_1 to E_2 , is that the level of output returns to potential GDP, where it started. There is, however, downward pressure on the price level. Thus, in the neoclassical view, changes in aggregate demand can have a short-run impact on output and on unemployment—but only a short-run impact. In the long run, when wages and prices are flexible, potential GDP and aggregate supply determine the size of real GDP.

How Fast Is the Speed of Macroeconomic Adjustment?

How long does it take for wages and prices to adjust, and for the economy to rebound back to its potential GDP? This subject is highly contentious. Keynesian economists argue that if the adjustment from recession to potential GDP takes a very long time, then neoclassical theory may be more

hypothetical than practical. In response to those immortal words of John Maynard Keynes, “In the long run we are all dead,” neoclassical economists respond that even if the adjustment takes as long as, say, ten years the neoclassical perspective remains of central importance in understanding the economy.

One subset of neoclassical economists holds that the adjustment of wages and prices in the macroeconomy might be quite rapid indeed. The theory of **rational expectations** holds that people form the most accurate possible expectations about the future that they can, using all information available to them. In an economy where most people have rational expectations, economic adjustments may happen very quickly.

To understand how rational expectations may affect the speed of price adjustments, think about a situation in the real estate market. Imagine that several events seem likely to push up the value of homes in the neighborhood. Perhaps a local employer announces that it is going to hire many more people or the city announces that it is going to build a local park or a library in that neighborhood. The theory of rational expectations points out that even though none of the changes will happen immediately, home prices in the neighborhood will rise immediately, because the expectation that homes will be worth more in the future will lead buyers to be willing to pay more in the present. The amount of the immediate increase in home prices will depend on how likely it seems that the announcements about the future will actually happen and on how distant the local jobs and neighborhood improvements are in the future. The key point is that, because of rational expectations, prices do not wait on events, but adjust immediately.

At a macroeconomic level, the theory of rational expectations points out that if the aggregate supply curve is vertical over time, then people should rationally expect this pattern. When a shift in aggregate demand occurs, people and businesses with rational expectations will know that its impact on output and employment will be temporary, while its impact on the price level will be permanent. If firms and workers perceive the outcome of the process in advance, and if all firms and workers know that everyone else is perceiving the process in the same way, then they have no incentive to go

through an extended series of short-run scenarios, like a firm first hiring more people when aggregate demand shifts out and then firing those same people when aggregate supply shifts back. Instead, everyone will recognize where this process is heading—toward a change in the price level—and then will act on that expectation. In this scenario, the expected long-run change in the price level may happen very quickly, without a drawn-out zigzag of output and employment first moving one way and then the other.

The theory that people and firms have rational expectations can be a useful simplification, but as a statement about how people and businesses actually behave, the assumption seems too strong. After all, many people and firms are not especially well informed, either about what is happening in the economy or about how the economy works. An alternate assumption is that people and firms act with **adaptive expectations**: they look at past experience and gradually adapt their beliefs and behavior as circumstances change, but are not perfect synthesizers of information and accurate predictors of the future in the sense of rational expectations theory. If most people and businesses have some form of adaptive expectations, then the adjustment from the short run and long run will be traced out in incremental steps that occur over time.

The empirical evidence on the speed of macroeconomic adjustment of prices and wages is not clear-cut. Indeed, the speed of macroeconomic adjustment probably varies among different countries and time periods. A reasonable guess is that the initial short-run effect of a shift in aggregate demand might last two to five years, before the adjustments in wages and prices cause the economy to adjust back to potential GDP. Thus, one might think of the short run for applying Keynesian analysis as time periods less than two to five years, and the long run for applying neoclassical analysis as longer than five years. For practical purposes, this guideline is frustratingly imprecise, but when analyzing a complex social mechanism like an economy as it evolves over time, some imprecision seems unavoidable.

Key Concepts and Summary

Neoclassical perspective argues that, in the long run, the economy will adjust back to its potential GDP level of output through flexible price

levels. Thus, the neoclassical perspective views the long-run AS curve as vertical. A rational expectations perspective argues that people have excellent information about economic events and how the economy works and that, as a result, price and other economic adjustments will happen very quickly. In adaptive expectations theory, people have limited information about economic information and how the economy works, and so price and other economic adjustments can be slow.

Self-Check Question

Exercise:

Problem:

Do rational expectations tend to look back at past experience while adaptive expectations look ahead to the future? Explain your answer.

Solution:

No, this statement is false. It would be more accurate to say that rational expectations seek to predict the future as accurately as possible, using all of past experience as a guide. Adaptive expectations are largely backward looking; that is, they adapt as experience accumulates, but without attempting to look forward.

Review Questions

Exercise:

Problem:

Does neoclassical economics focus on the long term or the short term? Explain your answer.

Exercise:

Problem:

Does neoclassical economics view prices and wages as sticky or flexible? Why?

Exercise:

Problem:

What shape is the long-run aggregate supply curve? Why does it have this shape?

Exercise:

Problem:

What is the difference between rational expectations and adaptive expectations?

Exercise:

Problem:

A neoclassical economist and a Keynesian economist are studying the economy of Vineland. It appears that Vineland is beginning to experience a mild recession with a decrease in aggregate demand. Which of these two economists would likely advocate that the government of Vineland take active measures to reverse this decline in aggregate demand? Why?

Critical Thinking Questions

Exercise:

Problem:

If most people have rational expectations, how long will recessions last?

Exercise:

Problem:

Explain why the neoclassical economists believe that nothing much needs to be done about unemployment. Do you agree or disagree? Explain.

Problems**Exercise:**

Problem: Use [\[link\]](#) to answer the following questions.

Price Level	Aggregate Supply	Aggregate Demand
90	3,000	3,500
95	3,000	3,000
100	3,000	2,500
105	3,000	2,200
110	3,000	2,100

- Sketch an aggregate supply and aggregate demand diagram.
- What is the equilibrium output and price level?
- If aggregate demand shifts right, what is equilibrium output?
- If aggregate demand shifts left, what is equilibrium output?
- In this scenario, would you suggest using aggregate demand to alter the level of output or to control any inflationary increases in

the price level?

References

Lumina Foundation. 2014. "A Stronger Nation Through Higher Education." Accessed March 4, 2015.

http://www.luminafoundation.org/publications/A_stronger_nation_through_higher_education-2014.pdf.

The National Bureau of Economic Research. <http://www.nber.org/>.

U.S. Department of Commerce: United States Census Bureau. "The 2012 Statistical Abstract."

<http://www.census.gov/compendia/statab/cats/education.html>.

U.S. Department of the Treasury. "TARP Programs." Last modified December 12, 2013. <http://www.treasury.gov/initiatives/financial-stability/TARP-Programs/Pages/default.aspx>.

United States Government. "Recovery.gov: Track the Money." Last modified October 30, 2013. <http://www.recovery.gov/Pages/default.aspx>.

Glossary

adaptive expectations

the theory that people look at past experience and gradually adapt their beliefs and behavior as circumstances change

neoclassical perspective

the philosophy that, in the long run, the business cycle will fluctuate around the potential, or full-employment, level of output

physical capital per person

the amount and kind of machinery and equipment available to help a person produce a good or service

rational expectations

the theory that people form the most accurate possible expectations about the future that they can, using all information available to them

The Policy Implications of the Neoclassical Perspective

By the end of this section, you will be able to:

- Discuss why and how inflation expectations are measured
- Analyze the impacts of fiscal policy and monetary policy on aggregate supply and aggregate demand
- Explain the neoclassical Phillips curve, noting its tradeoff between inflation and unemployment
- Identify clear distinctions between neoclassical economics and Keynesian economics

To understand the policy recommendations of the neoclassical economists, it helps to start with the Keynesian perspective. Suppose a decrease in aggregate demand causes the economy to go into recession with high unemployment. The Keynesian response would be to use government policy to stimulate aggregate demand and eliminate the recessionary gap. The neoclassical economists believe that the Keynesian response, while perhaps well intentioned, will not have a good outcome for reasons we will discuss shortly. Since the neoclassical economists believe that the economy will correct itself over time, the only advantage of a Keynesian stabilization policy would be to speed up the process and minimize the time that the unemployed are out of work. Is that the likely outcome?

Keynesian macroeconomic policy requires some optimism about the ability of the government to recognize a situation of too little or too much aggregate demand, and to adjust aggregate demand accordingly with the right level of changes in taxes or spending, all enacted in a timely fashion. After all, neoclassical economists argue, it takes government statisticians months to produce even preliminary estimates of GDP so that politicians know whether a recession is occurring—and those preliminary estimates may be revised substantially later. Moreover, there is the question of timely action. The political process can take more months to enact a tax cut or a spending increase; the amount of those tax or spending changes may be determined as much by political considerations as economic ones; and then the economy will take still more months to put changes in aggregate demand into effect through spending and production. When all of these time lags and political realities are considered, active fiscal policy may fail to

address the current problem, and could even make the future economy worse. The average U.S. post-World War II recession has lasted only about a year. By the time government policy kicks in, the recession will likely be over. As a consequence, the only result of government fine-tuning will be to stimulate the economy when it is already recovering (or to contract the economy when it is already falling). In other words, an active macroeconomic policy is likely to exacerbate the cycles rather than dampen them. Indeed, some neoclassical economists believe a large part of the business cycles we observe are due to flawed government policy. To learn about this issue further, read the following Clear It Up feature.

Note:

Why and how are inflation expectations measured?

People take expectations about inflation into consideration every time they make a major purchase, such as a house or a car. As inflation fluctuates, so too does the nominal interest rate on loans to buy these goods. The nominal interest rate is comprised of the real rate, plus an **expected inflation** factor. Expected inflation also tells economists about how the public views the direction of the economy. Suppose the public expects inflation to increase. This could be the result of positive demand shock due to an expanding economy and increasing aggregate demand. It could also be the result of a negative supply shock, perhaps from rising energy prices, and decreasing aggregate supply. In either case, the public may expect the central bank to engage in contractionary monetary policy to reduce inflation, and this policy results in higher interest rates. If, on the other hand, inflation is expected to decrease, the public may anticipate a recession. In turn, the public may expect expansionary monetary policy, and the lowering of interest rates, in the short run. By monitoring expected inflation, economists garner information about the effectiveness of macroeconomic policies. Additionally, monitoring expected inflation allows for projecting the direction of real interest rates that isolate for the effect of inflation. This information is necessary for making decisions about financing investments. Expectations about inflation may seem like a highly theoretical concept, but, in fact, inflation expectations are measured by the Federal Reserve Bank based upon early research conducted by Joseph Livingston, a

financial journalist for the *Philadelphia Inquirer*. In 1946, he started a twice-a-year survey of economists about their expectations of inflation. After Livingston's death in 1969, the survey was continued by the Federal Reserve Bank and other economic research agencies such as the Survey Research Center at the University of Michigan, the American Statistical Association, and the National Bureau of Economic Research. Current research by the Federal Reserve compares these expectations to actual inflation that has occurred, and the results, so far, are mixed. Economists' forecasts, however, have become notably more accurate in the last few decades. Economists are actively researching how expectations of inflation and other economic variables are formed and changed.

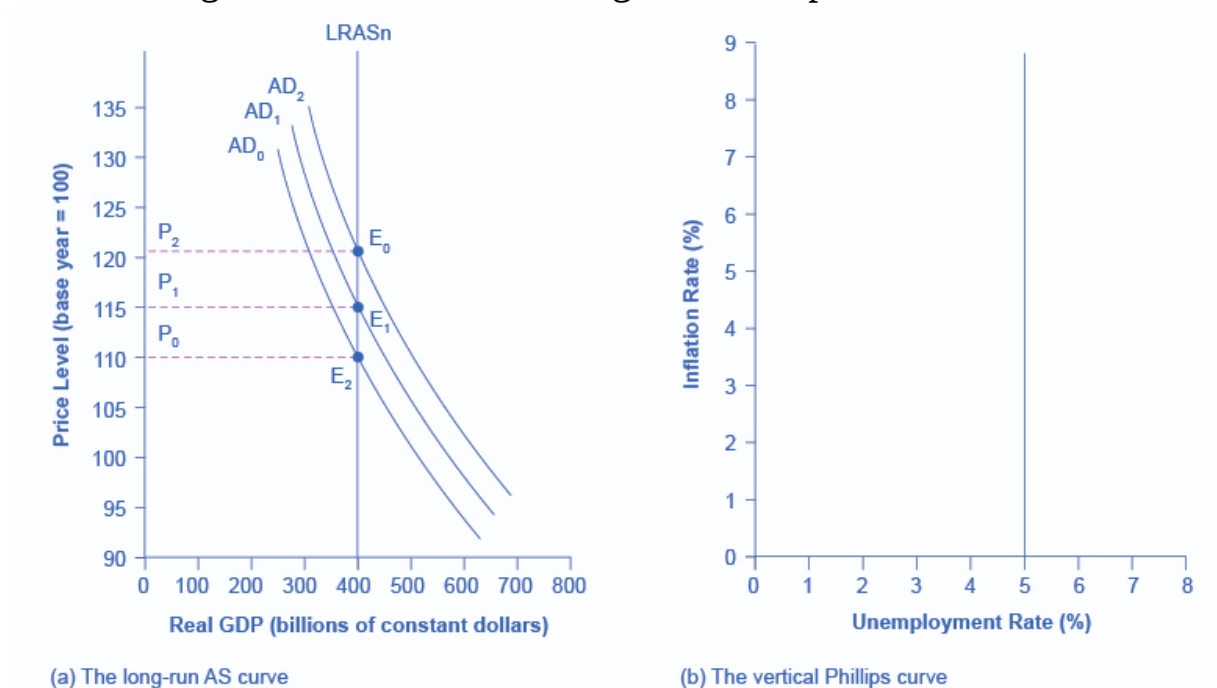
Note: Visit this [website](#) to read “The Federal Reserve Bank of Cleveland’s Economic Commentary: A New Approach to Gauging Inflation Expectations” by Joseph G. Haubrich for more information about how expected inflation is forecast.

The Neoclassical Phillips Curve Tradeoff

[The Keynesian Perspective](#) introduced the Phillips curve and explained how it is derived from the aggregate supply curve. The short run upward sloping aggregate supply curve implies a downward sloping Phillips curve; thus, there is a tradeoff between inflation and unemployment in the short run. By contrast, a neoclassical long-run aggregate supply curve will imply a vertical shape for the Phillips curve, indicating no long run tradeoff between inflation and unemployment. [\[link\]](#) (a) shows the vertical AS curve, with three different levels of aggregate demand, resulting in three different equilibria, at three different price levels. At every point along that vertical AS curve, potential GDP and the rate of unemployment remains the same. Assume that for this economy, the natural rate of unemployment is 5%. As a result, the long-run Phillips curve relationship, shown in [\[link\]](#) (b), is a vertical line, rising up from 5% unemployment, at any level of inflation.

Read the following Work It Out feature for additional information on how to interpret inflation and unemployment rates.

From a Long-Run AS Curve to a Long-Run Phillips Curve



(a) The long-run AS curve

(b) The vertical Phillips curve

(a) With a vertical LRAS curve, shifts in aggregate demand do not alter the level of output but do lead to changes in the price level. Because output is unchanged between the equilibria E₀, E₁, and E₂, all unemployment in this economy will be due to the natural rate of unemployment. (b) If the natural rate of unemployment is 5%, then the Phillips curve will be vertical. That is, regardless of changes in the price level, the unemployment rate remains at 5%.

Note:

Tracking Inflation and Unemployment Rates

Suppose that you have collected data for years on the rates of inflation and unemployment and recorded them in a table, such as [\[link\]](#). How do you interpret that information?

Year	Inflation Rate	Unemployment Rate
1970	2%	4%
1975	3%	3%
1980	2%	4%
1985	1%	6%
1990	1%	4%
1995	4%	2%
2000	5%	4%

Step 1. Plot the data points in a graph with inflation rate on the vertical axis and unemployment rate on the horizontal axis. Your graph will appear similar to [\[link\]](#).

Inflation Rates



Step 2. What patterns do you see in the data? You should notice that there are years when unemployment falls but inflation rises, and other years where unemployment rises and inflation falls.

Step 3. Can you determine the natural rate of unemployment from the data or from the graph? As you analyze the graph, it appears that the natural rate of unemployment lies at 4%; this is the rate that the economy appears

to adjust back to after an apparent change in the economy. For example, in 1975 the economy appeared to have an increase in aggregate demand; the unemployment rate fell to 3% but inflation increased from 2% to 3%. By 1980, the economy had adjusted back to 4% unemployment and the inflation rate had returned to 2%. In 1985, the economy looks to have suffered a recession as unemployment rose to 6% and inflation fell to 1%. This would be consistent with a decrease in aggregate demand. By 1990, the economy recovered back to 4% unemployment, but at a lower inflation rate of 1%. In 1995 the economy again rebounded and unemployment fell to 2%, but inflation increased to 4%, which is consistent with a large increase in aggregate demand. The economy adjusted back to 4% unemployment but at a higher rate of inflation of 5%. Then in 2000, both unemployment and inflation increased to 5% and 4%, respectively.

Step 4. Do you see the Phillips curve(s) in the data? If we trace the downward sloping trend of data points, we could see a short-run Phillips curve that exhibits the inverse tradeoff between higher unemployment and lower inflation rates. If we trace the vertical line of data points, we could see a long-run Phillips curve at the 4% natural rate of unemployment.

The unemployment rate on the long-run Phillips curve will be the natural rate of unemployment. A small inflationary increase in the price level from AD_0 to AD_1 will have the same natural rate of unemployment as a larger inflationary increase in the price level from AD_0 to AD_2 . The macroeconomic equilibrium along the vertical aggregate supply curve can occur at a variety of different price levels, and the natural rate of unemployment can be consistent with all different rates of inflation. The great economist Milton Friedman (1912–2006) summed up the neoclassical view of the long-term Phillips curve tradeoff in a 1967 speech: “[T]here is always a temporary trade-off between inflation and unemployment; there is no permanent trade-off.”

In the Keynesian perspective, the primary focus is on getting the level of aggregate demand right in relationship to an upward-sloping aggregate supply curve. That is, AD should be adjusted so that the economy produces at its potential GDP, not so low that cyclical unemployment results and not

so high that inflation results. In the neoclassical perspective, aggregate supply will determine output at potential GDP, unemployment is determined by the natural rate of unemployment churned out by the forces of supply and demand in the labor market, and shifts in aggregate demand are the primary determinant of changes in the price level.

Fighting Unemployment or Inflation?

As explained in [Unemployment](#), unemployment can be divided into two categories: cyclical unemployment and the natural rate of unemployment, which is the sum of frictional and structural unemployment. Cyclical unemployment results from fluctuations in the business cycle and is created when the economy is producing below potential GDP—giving potential employers less incentive to hire. When the economy is producing at potential GDP, cyclical unemployment will be zero. Because of the dynamics of the labor market, in which people are always entering or exiting the labor force, the unemployment rate never falls to 0%, not even when the economy is producing at or even slightly above potential GDP. Probably the best we can hope for is for the number of job vacancies to equal the number of job seekers. We know that it takes time for job seekers and employers to find each other, and this time is the cause of frictional unemployment. Most economists do not consider frictional unemployment to be a “bad” thing. After all, there will always be workers who are unemployed while looking for a job that is a better match for their skills. There will always be employers that have an open position, while looking for a worker that is a better match for the job. Ideally, these matches happen quickly, but even when the economy is very strong there will be some natural unemployment and this is what is measured by the natural rate of unemployment.

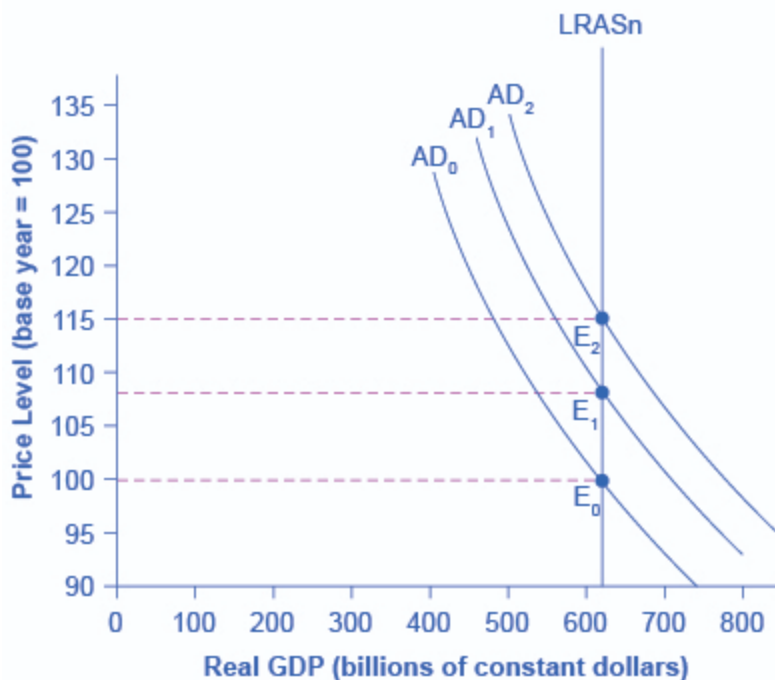
The neoclassical view of unemployment tends to focus attention away from the problem of cyclical unemployment—that is, unemployment caused by recession—while putting more attention on the issue of the rates of unemployment that prevail even when the economy is operating at potential GDP. To put it another way, the neoclassical view of unemployment tends to focus on how public policy can be adjusted to reduce the natural rate of unemployment. Such policy changes might involve redesigning

unemployment and welfare programs so that they support those in need, but also offer greater encouragement for job-hunting. It might involve redesigning business rules with an eye to whether they are unintentionally discouraging businesses from taking on new employees. It might involve building institutions to improve the flow of information about jobs and the mobility of workers, to help bring workers and employers together more quickly. For those workers who find that their skills are permanently no longer in demand (for example, the structurally unemployed), policy can be designed to provide opportunities for retraining so that these workers can reenter the labor force and seek employment.

Neoclassical economists will not tend to see aggregate demand as a useful tool for reducing unemployment; after all, if economic output is determined by a vertical aggregate supply curve, then aggregate demand has no long-run effect on unemployment. Instead, neoclassical economists believe that aggregate demand should be allowed to expand only to match the gradual shifts of aggregate supply to the right—keeping the price level much the same and inflationary pressures low.

If aggregate demand rises rapidly in the neoclassical model, in the long run it leads only to inflationary pressures. [\[link\]](#) shows a vertical LRAS curve and three different levels of aggregate demand, rising from AD_0 to AD_1 to AD_2 . As the macroeconomic equilibrium rises from E_0 to E_1 to E_2 , the price level rises, but real GDP does not budge; nor does the rate of unemployment, which adjusts to its natural rate. Conversely, reducing inflation has no long-term costs, either. Think about [\[link\]](#) in reverse, as the aggregate demand curve shifts from AD_2 to AD_1 to AD_0 , and the equilibrium moves from E_2 to E_1 to E_0 . During this process, the price level falls, but, in the long run, neither real GDP nor the natural rate of unemployment is changed.

How Aggregate Demand Determines the Price Level in the Long Run



As aggregate demand shifts to the right, from AD_0 to AD_1 to AD_2 , real GDP in this economy and the level of unemployment do not change. However, there is inflationary pressure for a higher price level as the equilibrium changes from E_0 to E_1 to E_2 .

Fighting Recession or Encouraging Long-Term Growth?

Neoclassical economists believe that the economy will rebound out of a recession or eventually contract during an expansion because prices and wage rates are flexible and will adjust either upward or downward to restore the economy to its potential GDP. Thus, the key policy question for neoclassicals is how to promote growth of potential GDP. We know that economic growth ultimately depends on the growth rate of long-term productivity. Productivity measures how effective inputs are at producing outputs. We know that U.S. productivity has grown on average about 2%

per year. That means that the same amount of inputs produce 2% more output than the year before. We also know that productivity growth varies a great deal in the short term due to cyclical factors. It also varies somewhat in the long term. From 1953–1972, U.S. labor productivity (as measured by output per hour in the business sector) grew at 3.2% per year. From 1973–1992, productivity growth declined significantly to 1.8% per year. Then, from 1993–2014, productivity growth increased slightly to 2% per year. The neoclassical economists believe the underpinnings of long-run productivity growth to be an economy’s investments in human capital, physical capital, and technology, operating together in a market-oriented environment that rewards innovation. Promotion of these factors is what government policy should focus on.

Summary of Neoclassical Macroeconomic Policy Recommendations

Let’s summarize what neoclassical economists recommend for macroeconomic policy. Neoclassical economists do not believe in “fine-tuning” the economy. They believe that economic growth is fostered by a stable economic environment with a low rate of inflation. Similarly, tax rates should be low and unchanging. In this environment, private economic agents can make the best possible investment decisions, which will lead to optimal investment in physical and human capital as well as research and development to promote improvements in technology.

Summary of Neoclassical Economics versus Keynesian Economics

[\[link\]](#) summarizes the key differences between the two schools of thought.

Summary	Neoclassical Economics	Keynesian Economics
Focus: long-term or short term	Long-term	Short-term
Prices and wages: sticky or flexible?	Flexible	Sticky
Economic output: Primarily determined by aggregate demand or aggregate supply?	Aggregate supply	Aggregate demand
Aggregate supply: vertical or upward-sloping?	Vertical	Upward-sloping
Phillips curve vertical or downward-sloping	Vertical	Downward sloping
Is aggregate demand a useful tool for controlling inflation?	Yes	Yes
What should be the primary area of policy emphasis for reducing unemployment?	Reform labor market institutions to reduce natural rate of unemployment	Increase aggregate demand to eliminate cyclical unemployment

Summary	Neoclassical Economics	Keynesian Economics
Is aggregate demand a useful tool for ending recession?	At best, only in the short-run temporary sense, but may just increase inflation instead	Yes

Neoclassical versus Keynesian Economics

Key Concepts and Summary

Neoclassical economists tend to put relatively more emphasis on long-term growth than on fighting recession, because they believe that recessions will fade in a few years and long-term growth will ultimately determine the standard of living. They tend to focus more on reducing the natural rate of unemployment caused by economic institutions and government policies than the cyclical unemployment caused by recession.

Neoclassical economists also see no social benefit to inflation. With an upward-sloping Keynesian AS curve, inflation can arise because an economy is approaching full employment. With a vertical long-run neoclassical AS curve, inflation does not accompany any rise in output. If aggregate supply is vertical, then aggregate demand does not affect the quantity of output. Instead, aggregate demand can only cause inflationary changes in the price level. A vertical aggregate supply curve, where the quantity of output is consistent with many different price levels, also implies a vertical Phillips curve.

Self-Check Questions

Exercise:

Problem:

Legislation proposes that the government should use macroeconomic policy to achieve an unemployment rate of zero percent, by increasing aggregate demand for as much and as long as necessary to accomplish this goal. From a neoclassical perspective, how will this policy affect output and the price level in the short run and in the long run? Sketch an aggregate demand/aggregate supply diagram to illustrate your answer. *Hint: revisit [\[link\]](#).*

Solution:

An unemployment rate of zero percent is presumably well below the rate that is consistent with potential GDP and with the natural rate of unemployment. As a result, this policy would be attempting to push AD out to the right.

In the short run, it is possible to have unemployment slightly below the natural rate for a time, at a price of higher inflation, as shown by the movement from E_0 to E_1 along the short-run AS curve. However, over time the extremely low unemployment rates will tend to cause wages to be bid up, and shift the short-run AS curve back to the left. The result would be a higher price level, but an economy still at potential GDP and the natural rate of unemployment, as determined by the long-run AS curve. If the government continues this policy, it will continually be pushing the price level higher and higher, but it will not be able to achieve its goal of zero percent unemployment, because that goal is inconsistent with market forces.

Exercise:**Problem:**

Would it make sense to argue that rational expectations economics is an extreme version of neoclassical economics? Explain.

Solution:

The statement is accurate. Rational expectations can be thought of as a version of neoclassical economics because it argues that potential GDP and the rate of unemployment are shaped by market forces as wages and prices adjust. However, it is an “extreme” version because it argues that this adjustment takes place very quickly. Other theories, like adaptive expectations, suggest that adjustment to the neoclassical outcome takes a few years.

Review Questions

Exercise:

Problem:

Do neoclassical economists tend to focus more on long term economic growth or on recessions? Explain briefly.

Exercise:

Problem:

Do neoclassical economists tend to focus more on cyclical unemployment or on inflation? Explain briefly.

Exercise:

Problem:

Do neoclassical economists see a value in tolerating a little more inflation if it brings additional economic output? Explain your answer.

Exercise:

Problem:

If aggregate supply is vertical, what role does aggregate demand play in determining output? In determining the price level?

Exercise:

Problem:

What is the shape of the neoclassical long-run Phillips curve? What assumptions are made that lead to this shape?

Critical Thinking Question**Exercise:****Problem:**

The American Recovery and Reinvestment Act was criticized by economists from all theoretical persuasions. The “Stimulus Package” was arguably a Keynesian measure so why would a Keynesian economist be critical of it? Why would neoclassical economists be critical?

References

American Statistical Association. “ASA Headlines.”
<http://www.amstat.org/>.

Haubrich, Joseph G., George Pennacchi, and Peter Ritchken. “Working Paper 11-07: Inflation Expectations, Real Rates, and Risk Premia: Evidence from Inflation Swaps.” *Federal Reserve Bank of Cleveland*. Last modified March 2011.
<http://www.clevelandfed.org/research/workpaper/2011/wp1107.pdf>.

University of Michigan: Institute for Social Research. “Survey Research Center.” <http://www.src.isr.umich.edu/>.

Glossary

expected inflation

a future rate of inflation that consumers and firms build into current decision making

Balancing Keynesian and Neoclassical Models

By the end of this section, you will be able to:

- Evaluate how neoclassical economists and Keynesian economists react to recessions
- Analyze the interrelationship between the neoclassical and Keynesian economic models

Finding the balance between Keynesian and Neoclassical models can be compared to the challenge of riding two horses simultaneously. When a circus performer stands on two horses, with a foot on each one, much of the excitement for the viewer lies in contemplating the gap between the two. As modern macroeconomists ride into the future on two horses—with one foot on the short-term Keynesian perspective and one foot on the long-term neoclassical perspective—the balancing act may look uncomfortable, but there does not seem to be any way to avoid it. Each approach, Keynesian and neoclassical, has its strengths and weaknesses.

The short-term Keynesian model, built on the importance of aggregate demand as a cause of business cycles and a degree of wage and price rigidity, does a sound job of explaining many recessions and why cyclical unemployment rises and falls. By focusing on the short-run adjustments of aggregate demand, Keynesian economics risks overlooking the long-term causes of economic growth or the natural rate of unemployment that exists even when the economy is producing at potential GDP.

The neoclassical model, with its emphasis on aggregate supply, focuses on the underlying determinants of output and employment in markets, and thus tends to put more emphasis on economic growth and how labor markets work. However, the neoclassical view is not especially helpful in explaining why unemployment moves up and down over short time horizons of a few years. Nor is the neoclassical model especially helpful when the economy is mired in an especially deep and long-lasting recession, like the Great Depression of the 1930s. Keynesian economics tends to view inflation as a price that might sometimes be paid for lower unemployment; neoclassical economics tends to view inflation as a cost that offers no offsetting gains in terms of lower unemployment.

Macroeconomics cannot, however, be summed up as an argument between one group of economists who are pure Keynesians and another group who are pure neoclassicists. Instead, many mainstream economists believe both the Keynesian and neoclassical perspectives. Robert Solow, the Nobel laureate in economics in 1987, described the dual approach in this way:

"At short time scales, I think, something sort of 'Keynesian' is a good approximation, and surely better than anything straight 'neoclassical.' At very long time scales, the interesting questions are best studied in a neoclassical framework, and attention to the Keynesian side of things would be a minor distraction. At the five-to-ten-year time scale, we have to piece things together as best we can, and look for a hybrid model that will do the job."

Many modern macroeconomists spend considerable time and energy trying to construct models that blend the most attractive aspects of the Keynesian and neoclassical approaches. It is possible to construct a somewhat complex mathematical model where aggregate demand and sticky wages and prices matter in the short run, but wages, prices, and aggregate supply adjust in the long run. However, creating an overall model that encompasses both short-term Keynesian and long-term neoclassical models is not easy.

Note:

Navigating Uncharted Waters

Were the policies implemented to stabilize the economy and financial markets during the Great Recession effective? Many economists from both the Keynesian and neoclassical schools have found that they were, although to varying degrees. Alan Blinder of Princeton University and Mark Zandi for Moody's Analytics found that, without fiscal policy, GDP decline would have been significantly more than its 3.3% in 2008 followed by its 0.1% decline in 2009. They also estimated that there would have been 8.5 million more job losses had the government not intervened in the market with the TARP to support the financial industry and key automakers General Motors and Chrysler. Federal Reserve Bank economists Carlos Carvalho, Stefano Eusip, and Christian Grisse found in their study, *Policy Initiatives in the Global Recession: What Did*

Forecasters Expect? that once policies were implemented, forecasters adapted their expectations to these policies. They were more likely to anticipate increases in investment due to lower interest rates brought on by monetary policy and increased economic growth resulting from fiscal policy.

The difficulty with evaluating the effectiveness of the stabilization policies that were taken in response to the Great Recession is that we will never know what would have happened had those policies not have been implemented. Surely some of the programs were more effective at creating and saving jobs, while other programs were less so. The final conclusion on the effectiveness of macroeconomic policies is still up for debate, and further study will no doubt consider the impact of these policies on the U.S. budget and deficit, as well as the value of the U.S. dollar in the financial market.

Key Concepts and Summary

The Keynesian perspective considers changes to aggregate demand to be the cause of business cycle fluctuations. Keynesians are likely to advocate that policy makers actively attempt to reverse recessionary and inflationary periods because they are not convinced that the self-correcting economy can easily return to full employment.

The neoclassical perspective places more emphasis on aggregate supply. The level of potential GDP is determined by long term productivity growth and that the economy typically will return to full employment after a change in aggregate demand. Skeptical of the effectiveness and timeliness of Keynesian policy, neoclassical economists are more likely to advocate a hands-off, or fairly limited, role for active stabilization policy.

While Keynesians would tend to advocate an acceptable tradeoff between inflation and unemployment when counteracting a recession, neoclassical economists argue that no such tradeoff exists; any short-term gains in lower unemployment will eventually vanish and the result of active policy will only be inflation.

Self-Check Question

Exercise:

Problem: Summarize the Keynesian and Neoclassical models.

Solution:

The short-term Keynesian model is built on the importance of aggregate demand as a cause of business cycles and a degree of wage and price rigidity, and thus does a sound job of explaining many recessions and why cyclical unemployment rises and falls. The neoclassical model emphasizes aggregate supply by focusing on the underlying determinants of output and employment in markets, and thus tends to put more emphasis on economic growth and how labor markets work.

Review Questions

Exercise:

Problem:

When the economy is experiencing a recession, why would a neoclassical economist be unlikely to argue for aggressive policy to stimulate aggregate demand and return the economy to full employment? Explain your answer.

Exercise:

Problem:

If the economy is suffering through a rampant inflationary period, would a Keynesian economist advocate for stabilization policy that involves higher taxes and higher interest rates? Explain your answer.

Critical Thinking Question

Exercise:**Problem:**

Is it a logical contradiction to be a neoclassical Keynesian? Explain.

Reference

Carvalho, Carlos, Stefano Eusepi, and Christian Grosse. "Policy Initiatives in the Global Recession: What Did Forecasters Expect?" *Federal Reserve Bank of New York: Current Issues in Economics and Finance*, 18, no. 2 (2012). http://www.newyorkfed.org/research/current_issues/ci18-2.pdf.

Introduction to Environmental Protection and Negative Externalities

class="introduction"

Environmental Debate

Across the country, countless people have protested, even risking arrest, against the Keystone XL Pipeline.
(Credit: modification of image by “NoKXL”/Flickr r Creative Commons)



Note:
Keystone XL

You might have heard about Keystone XL in the news. It is a pipeline system designed to bring oil from Canada to the refineries near the Gulf of Mexico, as well as to boost crude oil production in the United States. While a private company, TransCanada, will own the pipeline, U.S. government approval is required because of its size and location. The pipeline is being built in four phases, with the first two currently in operation, bringing oil from Alberta, Canada, east across Canada, south through the United States into Nebraska and Oklahoma, and northeast again to Illinois. The third and fourth phases of the project, known as Keystone XL, would create a pipeline southeast from Alberta straight to Nebraska, and then from Oklahoma to the Gulf of Mexico.

Sounds like a great idea, right? A pipeline that would move much needed crude oil to the Gulf refineries would increase oil production for manufacturing needs, reduce price pressure at the gas pump, and increase overall economic growth. Supporters argue that the pipeline is one of the safest pipelines built yet, and would reduce America's dependence on politically vulnerable Middle Eastern oil imports.

Not so fast, say its critics. The Keystone XL would be constructed over an enormous aquifer (one of the largest in the world) in the Midwest, and through an environmentally fragile area in Nebraska, causing great concern among environmentalists about possible destruction to the natural surroundings. They argue that leaks could taint valuable water sources and construction of the pipeline could disrupt and even harm indigenous species. Environmentalist groups have fought government approval of the proposed construction of the pipeline, and as of press time the pipeline projects remain stalled.

Of course, environmental concerns matter when discussing issues related to economic growth. But how much should they factor in? In the case of the pipeline, how do we know how much damage it would cause when we do not know how to put a value on the environment? Would the benefits of the pipeline outweigh the opportunity cost? The issue of how to balance economic progress with unintended effects on our planet is the subject of this chapter.

Note:

Introduction to Environmental Protection and Negative Externalities

In this chapter, you will learn about:

- The Economics of Pollution
- Command-and-Control Regulation
- Market-Oriented Environmental Tools
- The Benefits and Costs of U.S. Environmental Laws
- International Environmental Issues
- The Tradeoff between Economic Output and Environmental Protection

In 1969, the Cuyahoga River in Ohio was so polluted that it spontaneously burst into flame. Air pollution was so bad at that time that Chattanooga, Tennessee was a city where, as an article from Sports Illustrated put it: “the death rate from tuberculosis was double that of the rest of Tennessee and triple that of the rest of the United States, a city in which the filth in the air was so bad it melted nylon stockings off women’s legs, in which executives kept supplies of clean white shirts in their offices so they could change when a shirt became too gray to be presentable, in which headlights were turned on at high noon because the sun was eclipsed by the gunk in the sky.”

The problem of pollution arises for every economy in the world, whether high-income or low-income, and whether market-oriented or command-oriented. Every country needs to strike some balance between production and environmental quality. This chapter begins by discussing how firms may fail to take certain social costs, like pollution, into their planning if they do not need to pay these costs. Traditionally, policies for environmental protection have focused on governmental limits on how much of each pollutant could be emitted. While this approach has had some success, economists have suggested a range of more flexible, market-oriented policies that reduce pollution at a lower cost. We will consider both approaches, but first let’s see how economists frame and analyze these issues.

The Economics of Pollution

By the end of this section, you will be able to:

- Explain and give examples of positive and negative externalities
- Identify equilibrium price and quantity
- Evaluate how firms can contribute to market failure

From 1970 to 2012, the U.S. population increased by one-third and the size of the U.S. economy more than doubled. Since the 1970s, however, the United States, using a variety of anti-pollution policies, has made genuine progress against a number of pollutants. [\[link\]](#) lists the change in carbon dioxide emissions by users of energy (from residential to industrial) according to the U.S. Energy Information Administration (EIA). The table shows that emissions of certain key air pollutants declined substantially from 2007 to 2012; they dropped 730 million metric tons (MMT) a year—a 12% reduction. This seems to indicate that progress has been made in the United States in reducing overall carbon dioxide emissions, which cause greenhouse gases.

	Primary Fossil Fuels			Purchased Electric Power	Total Primary Fossil Fuels
End-use Sector	Coal	Petroleum	Natural Gas		
Residential	(0)	(14)	(31)	(134)	(179)
Commercial	(2)	(2)	(7)	(126)	(136)
Industrial	(40)	(62)	31	(118)	(191)
Transportation	0	(228)	5	(1)	(224)
Power	(464)	(36)	(122)	-	-
Change 2007–2012	(508)	(342)	121	(378)	(730)

U.S. Carbon Dioxide (CO₂) Emissions from Fossil Fuels Consumed 2007–2012, Million Metric Tons (MMT) per Year (Source: EIA Monthly Energy Review)

about 10% (about 10%) per year (Source: Environmental Energy Review)

Despite the gradual reduction in emissions from fossil fuels, many important environmental issues remain. Along with the still high levels of air and water pollution, other issues include hazardous waste disposal, destruction of wetlands and other wildlife habitats, and the impact on human health from pollution.

Externalities

Private markets, such as the cell phone industry, offer an efficient way to put buyers and sellers together and determine what goods are produced, how they are produced, and who gets them. The principle that voluntary exchange benefits both buyers and sellers is a fundamental building block of the economic way of thinking. But what happens when a voluntary exchange affects a third party who is neither the buyer nor the seller?

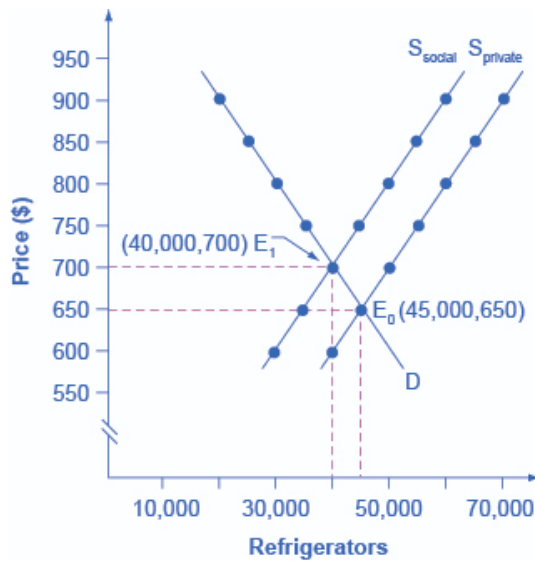
As an example, consider a concert producer who wants to build an outdoor arena that will host country music concerts a half-mile from your neighborhood. You will be able to hear these outdoor concerts while sitting on your back porch—or perhaps even in your dining room. In this case, the sellers and buyers of concert tickets may both be quite satisfied with their voluntary exchange, but you have no voice in their market transaction. The effect of a market exchange on a third party who is outside or “external” to the exchange is called an **externality**. Because externalities that occur in market transactions affect other parties beyond those involved, they are sometimes called **spillovers**.

Externalities can be negative or positive. If you hate country music, then having it waft into your house every night would be a **negative externality**. If you love country music, then what amounts to a series of free concerts would be a **positive externality**.

Pollution as a Negative Externality

Pollution is a negative externality. Economists illustrate the **social costs** of production with a demand and supply diagram. The social costs include the private costs of production incurred by the company and the external costs of pollution that are passed on to society. [\[link\]](#) shows the demand and supply for manufacturing refrigerators. The demand curve (D) shows the quantity demanded at each price. The supply curve (S_{private}) shows the quantity of refrigerators supplied by all the firms at each price if they are taking only their private costs into account and they are allowed to emit pollution at zero cost. The market equilibrium (E_0), where quantity supplied and quantity demanded are equal, is at a price of \$650 and a quantity of 45,000. This information is also reflected in the first three columns of [\[link\]](#).

Taking Social Costs into Account: A Supply Shift



If the firm takes only its own costs of production into account, then its supply curve will be $S_{private}$, and the market equilibrium will occur at E_0 . Accounting for additional external costs of \$100 for every unit produced, the firm's supply curve will be S_{social} . The new equilibrium will occur at E_1 .

Price	Quantity Demanded	Quantity Supplied before Considering Pollution Cost	Quantity Supplied after Considering Pollution Cost
\$600	50,000	40,000	30,000
\$650	45,000	45,000	35,000
\$700	40,000	50,000	40,000
\$750	35,000	55,000	45,000

Price	Quantity Demanded	Quantity Supplied before Considering Pollution Cost	Quantity Supplied after Considering Pollution Cost
\$800	30,000	60,000	50,000
\$850	25,000	65,000	55,000
\$900	20,000	70,000	60,000

A Supply Shift Caused by Pollution Costs

However, as a by-product of the metals, plastics, chemicals and energy that are used in manufacturing refrigerators, some pollution is created. Let's say that, if these pollutants were emitted into the air and water, they would create costs of \$100 per refrigerator produced. These costs might occur because of injuries to human health, property values, wildlife habitat, reduction of recreation possibilities, or because of other negative impacts. In a market with no anti-pollution restrictions, firms can dispose of certain wastes absolutely free. Now imagine that firms which produce refrigerators must factor in these external costs of pollution—that is, the firms have to consider not only the costs of labor and materials needed to make a refrigerator, but also the broader costs to society of injuries to health and other values caused by pollution. If the firm is required to pay \$100 for the **additional external costs** of pollution each time it produces a refrigerator, production becomes more costly and the entire supply curve shifts up by \$100.

As illustrated in the fourth column of [\[link\]](#) and in [\[link\]](#), the firm will need to receive a price of \$700 per refrigerator and produce a quantity of 40,000—and the firm's new supply curve will be S_{social} . The new equilibrium will occur at E_1 , taking the additional external costs of pollution into account results in a higher price, a lower quantity of production, and a lower quantity of pollution. The following Work It Out feature will walk you through an example, this time with musical accompaniment.

Note:

Identifying the Equilibrium Price and Quantity

[\[link\]](#) shows the supply and demand conditions for a firm that will play trumpets on the streets when requested. Output is measured as the number of songs played.

Price	Quantity Demanded	Quantity Supplied without paying the costs of the externality	Quantity Supplied after paying the costs of the externality
\$20	0	10	8
\$18	1	9	7
\$15	2.5	7.5	5.5
\$12	4	6	4
\$10	5	5	3
\$5	7.5	2.5	0.5

Supply and Demand Conditions for a Trumpet-Playing Firm

Step 1. Determine the negative externality in this situation. To do this, you must think about the situation described and consider all parties that might be impacted. A negative externality might be the increase in noise pollution in the area where the firm is playing.

Step 2. Identify the equilibrium price and quantity when only private costs are taken into account, and then when social costs are taken into account. Remember that equilibrium is where the quantity demanded is equal to the quantity supplied.

Step 3. Look down the columns to where the quantity demanded (the second column) is equal to the “quantity supplied without paying the costs of the externality” (the third column). Then refer to the first column of that row to determine the equilibrium price. In this case, the equilibrium price and quantity when only private costs are taken into account would be at a price of \$10 and a quantity of five.

Step 4. Identify the equilibrium price and quantity when the additional external costs are taken into account. Look down the columns of quantity demanded (the second column) and the “quantity supplied after paying the costs of the externality” (the fourth column) then refer to the first column of that row to determine the equilibrium price. In this case, the equilibrium will be at a price of \$12 and a quantity of four.

Step 5. Consider how taking the externality into account affects the equilibrium price and quantity. Do this by comparing the two equilibrium situations. If the firm is forced to pay its additional external costs, then production of trumpet songs becomes more costly, and the supply curve will shift up.

Remember that the supply curve is based on choices about production that firms make while looking at their marginal costs, while the demand curve is based on the benefits that individuals perceive while maximizing utility. If no externalities existed, private costs

would be the same as the costs to society as a whole, and private benefits would be the same as the benefits to society as a whole. Thus, if no externalities existed, the interaction of demand and supply will coordinate social costs and benefits.

However, when the externality of pollution exists, the supply curve no longer represents all social costs. Because externalities represent a case where markets no longer consider all social costs, but only some of them, economists commonly refer to externalities as an example of **market failure**. When there is market failure, the private market fails to achieve efficient output, because either firms do not account for all costs incurred in the production of output and/or consumers do not account for all benefits obtained (a positive externality). In the case of pollution, at the market output, social costs of production exceed social benefits to consumers, and the market produces too much of the product.

We can see a general lesson here. If firms were required to pay the social costs of pollution, they would create less pollution but produce less of the product and charge a higher price. In the next module, we will explore how governments require firms to take the social costs of pollution into account.

Key Concepts and Summary

Economic production can cause environmental damage. This tradeoff arises for all countries, whether high-income or low-income, and whether their economies are market-oriented or command-oriented.

An externality occurs when an exchange between a buyer and seller has an impact on a third party who is not part of the exchange. An externality, which is sometimes also called a spillover, can have a negative or a positive impact on the third party. If those parties imposing a negative externality on others had to take the broader social cost of their behavior into account, they would have an incentive to reduce the production of whatever is causing the negative externality. In the case of a positive externality, the third party is obtaining benefits from the exchange between a buyer and a seller, but they are not paying for these benefits. If this is the case, then markets would tend to under produce output because suppliers are not aware of the additional demand from others. If the parties that are generating benefits to others would be somehow compensated for these external benefits, they would have an incentive to increase production of whatever is causing the positive externality.

Self-Check Questions

Exercise:

Problem:

Identify the following situations as an example of a negative or a positive externality:

- a. You are a birder (bird watcher), and your neighbor has put up several birdhouses in the yard as well as planting trees and flowers that attract birds.
 - b. Your neighbor paints his house a hideous color.
 - c. Investments in private education raise your country's standard of living.
 - d. Trash dumped upstream flows downstream right past your home.
 - e. Your roommate is a smoker, but you are a nonsmoker.
-

Solution:

- a. positive externality
- b. negative externality
- c. positive externality
- d. negative externality
- e. negative externality

Exercise:

Problem:

Identify whether the market supply curve will shift right or left or will stay the same for the following:

- a. Firms in an industry are required to pay a fine for their emissions of carbon dioxide.
 - b. Companies are sued for polluting the water in a river.
 - c. Power plants in a specific city are not required to address the impact of their emissions on the quality of air.
 - d. Companies that use fracking to remove oil and gas from rock are required to clean up the damage.
-

Solution:

- a. supply shifts left
- b. supply shifts left
- c. supply stays the same
- d. supply shifts left

Exercise:

Problem:

For each of your answers to [\[link\]](#), will equilibrium price rise or fall or stay the same?

Solution:

- a. price will rise
- b. price will rise
- c. price stays the same
- d. price will rise.

Exercise:

Problem:

The supply and demand conditions for a manufacturing firm are given in [\[link\]](#). The third column represents a supply curve without taking the social cost of pollution into account. The fourth column represents the supply curve when the firm is required to take the social cost of pollution into account. Identify the equilibrium before the social cost of production is included and after the social cost of production is included.

Price	Quantity Demanded	Quantity Supplied without paying the cost of the pollution	Quantity Supplied after paying the cost of the pollution
\$10	450	400	250
\$15	440	440	290
\$20	430	480	330
\$25	420	520	370
\$30	410	560	410

Solution:

The original equilibrium (before the external social cost of pollution is taken into account) is where the private supply curve crosses the demand curve. This original equilibrium is at a price of \$15 and a quantity of 440. After taking into account the additional external cost of pollution, the production becomes more costly, and the supply curve shifts up. The new equilibrium will be at a price of \$30 and a quantity of 410.

Review Questions

Exercise:

Problem: What is an externality?

Exercise:

Problem:

Give an example of a positive externality and an example of a negative externality.

Exercise:

Problem: What is the difference between private costs and social costs?

Exercise:

Problem:

In a market without environmental regulations, will the supply curve for a firm take into account private costs, external costs, both, or neither? Explain.

Critical Thinking Questions

Exercise:

Problem:

Suppose you want to put a dollar value on the external costs of carbon emissions from a power plant. What information or data would you obtain to measure the external [not social] cost?

Problems

Exercise:

Problem:

Show the market for cigarettes in equilibrium, assuming that there are no laws banning smoking in public. Label the equilibrium private market price and quantity as P_m and Q_m . Add whatever is needed to the model to show the impact of the negative externality from second-hand smoking. (Hint: In this case it is the consumers, not the sellers, who are creating the negative externality.) Label the social optimal output and price as P_e and Q_e . On the graph, shade in the deadweight loss at the market output.

Exercise:

Problem:

Refer to [\[link\]](#). The externality created by the production of refrigerators was \$100. However, once both the private and additional external costs were taken into consideration, the market price increased by only \$50. If the external costs were \$100 why did the price only increase by \$50 when all costs were taken into account?

Exercise:**Problem:**

[\[link\]](#), shows the supply and demand conditions for a firm that will play trumpets on the streets when requested. Q_{s1} is the quantity supplied without social costs. Q_{s2} is the quantity supplied with social costs. What is the negative externality in this situation? Identify the equilibrium price and quantity when only private costs are taken into account, and then when social costs are taken into account. How does taking the externality into account affect the equilibrium price and quantity?

P	Qd	Q _{s1}	Q _{s2}
\$20	0	10	8
\$18	1	9	7
\$15	2.5	7.5	5.5
\$12	4	6	4
\$10	5	5	3
\$5	7.5	2.5	0.5

References

Johnson, Oscar William. "Back on Track: Earth Day Success Story; The Chattanooga Choo-Choo No Longer Spews Foul Air." *Sports Illustrated*. April 30, 1990.
<http://www.si.com/vault/1990/04/30/121923/back-on-track-earth-day-success-story-the-chattanooga-choo-choo-no-longer-spews-foul-air>.

U.S. Energy Information Administration. "Total Energy: Monthly Energy Review." *U.S. Department of Energy*. Accessed December 19, 2013.
<http://www.eia.gov/totalenergy/data/monthly/>.

Glossary

additional external cost

additional costs incurred by third parties outside the production process when a unit of output is produced

externality

a market exchange that affects a third party who is outside or "external" to the exchange; sometimes called a "spillover"

market failure

When the market on its own does not allocate resources efficiently in a way that balances social costs and benefits; externalities are one example of a market failure

negative externality

a situation where a third party, outside the transaction, suffers from a market transaction by others

positive externality

a situation where a third party, outside the transaction, benefits from a market transaction by others

social costs

costs that include both the private costs incurred by firms and also additional costs incurred by third parties outside the production process, like costs of pollution

spillover

see externality

Command-and-Control Regulation

By the end of this section, you will be able to:

- Explain command-and-control regulation
- Evaluate the effectiveness of command-and-control regulation

When the United States started passing comprehensive environmental laws in the late 1960s and early 1970s, a typical law specified how much pollution could be emitted out of a smokestack or a drainpipe and imposed penalties if that limit was exceeded. Other laws required the installation of certain equipment—for example, on automobile tailpipes or on smokestacks—to reduce pollution. These types of laws, which specify allowable quantities of pollution and which also may detail which pollution-control technologies must be used, fall under the category of **command-and-control regulation**. In effect, command-and-control regulation requires that firms increase their costs by installing anti-pollution equipment; firms are thus required to take the social costs of pollution into account.

Command-and-control regulation has been highly successful in protecting and cleaning up the U.S. environment. In 1970, the Environmental Protection Agency (EPA) was created to oversee all environmental laws. In the same year, the Clean Air Act was enacted to address air pollution. Just two years later, in 1972, Congress passed and the president signed the far-reaching Clean Water Act. These command-and-control environmental laws, and their amendments and updates, have been largely responsible for America's cleaner air and water in recent decades. However, economists have pointed out three difficulties with command-and-control environmental regulation.

First, command-and-control regulation offers no incentive to improve the quality of the environment beyond the standard set by a particular law. Once the command-and-control regulation has been satisfied, polluters have zero incentive to do better.

Second, command-and-control regulation is inflexible. It usually requires the same standard for all polluters, and often the same pollution-control technology as well. This means that command-and-control regulation draws

no distinctions between firms that would find it easy and inexpensive to meet the pollution standard—or to reduce pollution even further—and firms that might find it difficult and costly to meet the standard. Firms have no reason to rethink their production methods in fundamental ways that might reduce pollution even more and at lower cost.

Third, command-and-control regulations are written by legislators and the EPA, and so they are subject to compromises in the political process. Existing firms often argue (and lobby) that stricter environmental standards should not apply to them, only to new firms that wish to start production. Consequently, real-world environmental laws are full of fine print, loopholes, and exceptions.

Although critics accept the goal of reducing pollution, they question whether command-and-control regulation is the best way to design policy tools for accomplishing that goal. A different approach is the use of market-oriented tools, which are discussed in the next section.

Key Concepts and Summary

Command-and-control regulation sets specific limits for pollution emissions and/or specific pollution-control technologies that must be used. Although such regulations have helped to protect the environment, they have three shortcomings: they provide no incentive for going beyond the limits they set; they offer limited flexibility on where and how to reduce pollution; and they often have politically-motivated loopholes.

Self-Check Questions

Exercise:

Problem:

Consider two approaches to reducing emissions of CO₂ into the environment from manufacturing industries in the United States. In the first approach, the U.S. government makes it a policy to use only predetermined technologies. In the second approach, the U.S. government determines which technologies are cleaner and subsidizes their use. Of the two approaches, which is the command-and-control policy?

Solution:

The first policy is command-and-control because it is a requirement that applies to all producers.

Review Questions**Exercise:**

Problem: What is command-and-control environmental regulation?

Exercise:**Problem:**

What are the three problems that economists have noted with regard to command-and-control regulation?

Critical Thinking Questions**Exercise:****Problem:**

Would environmentalists favor command-and-control policies as a way to reduce pollution? Why or why not?

Exercise:

Problem:

Consider two ways of protecting elephants from poachers in African countries. In one approach, the government sets up enormous national parks that have sufficient habitat for elephants to thrive and forbids all local people to enter the parks or to injure either the elephants or their habitat in any way. In a second approach, the government sets up national parks and designates 10 villages around the edges of the park as official tourist centers that become places where tourists can stay and bases for guided tours inside the national park. Consider the different incentives of local villagers—who often are very poor—in each of these plans. Which plan seems more likely to help the elephant population?

Glossary

command-and-control regulation

laws that specify allowable quantities of pollution and that also may detail which pollution-control technologies must be used

Market-Oriented Environmental Tools

By the end of this section, you will be able to:

- Show how pollution charges impact firm decisions
- Suggest other laws and regulations that could fall under pollution charges
- Explain the significance of marketable permits and property rights
- Evaluate which policies are most appropriate for various situations

Market-oriented environmental policies create incentives to allow firms some flexibility in reducing pollution. The three main categories of market-oriented approaches to pollution control are pollution charges, marketable permits, and better-defined property rights. All of these policy tools, discussed below, address the shortcomings of command-and-control regulation—albeit in different ways.

Pollution Charges

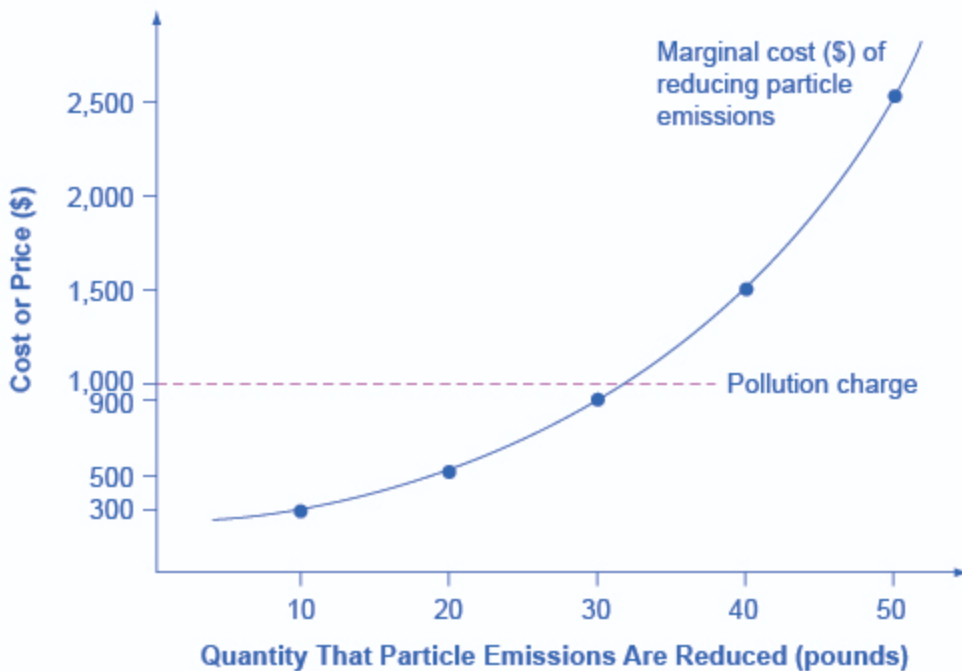
A **pollution charge** is a tax imposed on the quantity of pollution that a firm emits. A pollution charge gives a profit-maximizing firm an incentive to figure out ways to reduce its emissions—as long as the marginal cost of reducing the emissions is less than the tax.

For example, consider a small firm that emits 50 pounds per year of small particles, such as soot, into the air. Particulate matter, as it is called, causes respiratory illnesses and also imposes costs on firms and individuals.

[\[link\]](#) illustrates the marginal costs that a firm faces in reducing pollution. The marginal cost of pollution reduction, like most most marginal cost curves increases with output, at least in the short run. Reducing the first 10 pounds of particulate emissions costs the firm \$300. Reducing the second 10 pounds would cost \$500; reducing the third ten pounds would cost \$900; reducing the fourth 10 pounds would cost \$1,500; and the fifth 10 pounds would cost \$2,500. This pattern for the costs of reducing pollution is common, because the firm can use the cheapest and easiest method to make

initial reductions in pollution, but additional reductions in pollution become more expensive.

A Pollution Charge



If a pollution charge is set equal to \$1,000, then the firm will have an incentive to reduce pollution by 30 pounds because the \$900 cost of these reductions would be less than the cost of paying the pollution charge.

Imagine the firm now faces a pollution tax of \$1,000 for every 10 pounds of particulates emitted. The firm has the choice of either polluting and paying the tax, or reducing the amount of particulates they emit and paying the cost of abatement as shown in the figure. How much will the firm pollute and how much will the firm abate? The first 10 pounds would cost the firm \$300 to abate. This is substantially less than the \$1,000 tax, so they will choose to abate. The second 10 pounds would cost \$500 to abate, which is still less than the tax, so they will choose to abate. The third 10 pounds would cost \$900 to abate, which is slightly less than the \$1,000 tax. The fourth 10 pounds would cost \$1,500, which is much more costly than

paying the tax. As a result, the firm will decide to reduce pollutants by 30 pounds, because the marginal cost of reducing pollution by this amount is less than the pollution tax. With a tax of \$1,000, the firm has no incentive to reduce pollution more than 30 pounds.

A firm that has to pay a pollution tax will have an incentive to figure out the least expensive technologies for reducing pollution. Firms that can reduce pollution cheaply and easily will do so to minimize their pollution taxes, whereas firms that will incur high costs for reducing pollution will end up paying the pollution tax instead. If the pollution tax applies to every source of pollution, then no special favoritism or loopholes are created for politically well-connected producers.

For an example of a pollution charge at the household level, consider two ways of charging for garbage collection. One method is to have a flat fee per household, no matter how much garbage a household produces. An alternative approach is to have several levels of fees, depending on how much garbage the household produces—and to offer lower or free charges for recyclable materials. As of 2006 (latest statistics available), the EPA had recorded over 7,000 communities that have implemented “pay as you throw” programs. When people have a financial incentive to put out less garbage and to increase recycling, they find ways of doing so.

Note:

Visit this [website](#) to learn more about pay-as-you-throw programs, including viewing a map and a table that shows the number of communities using this program in each state.

A number of environmental policies are really pollution charges, although they often do not travel under that name. For example, the federal government and many state governments impose taxes on gasoline. We can view this tax as a charge on the air pollution that cars generate as well as a source of funding for maintaining roads. Indeed, gasoline taxes are far higher in most other countries than in the United States.

Similarly, the refundable charge of five or 10 cents that only 10 states have for returning recyclable cans and bottles works like a pollution tax that provides an incentive to avoid littering or throwing bottles in the trash. Compared with command-and-control regulation, a pollution tax reduces pollution in a more flexible and cost-effective way.

Marketable Permits

When a city or state government sets up a **marketable permit program** (e.g. cap-and-trade), it must start by determining the overall quantity of pollution it will allow as it tries to meet national pollution standards. Then, a number of permits allowing only this quantity of pollution are divided among the firms that emit that pollutant. These permits to pollute can be sold or given to firms free.

Now, add two more conditions. Imagine that these permits are designed to reduce total emissions over time. For example, a permit may allow emission of 10 units of pollution one year, but only nine units the next year, then eight units the year after that, and so on down to some lower level. In addition, imagine that these are marketable permits, meaning that firms can buy and sell them.

To see how marketable permits can work to reduce pollution, consider the four firms listed in [\[link\]](#). The table shows current emissions of lead from each firm. At the start of the marketable permit program, each firm receives permits to allow this level of pollution. However, these permits are shrinkable, and next year the permits allow the firms to emit only half as much pollution. Let's say that in a year, Firm Gamma finds it easy and cheap to reduce emissions from 600 tons of lead to 200 tons, which means that it has permits that it is not using that allow emitting 100 tons of lead. Firm Beta reduces its lead pollution from 400 tons to 200 tons, so it does not need to buy any permits, and it does not have any extra permits to sell. However, although Firm Alpha can easily reduce pollution from 200 tons to 150 tons, it finds that it is cheaper to purchase permits from Gamma rather than to reduce its own emissions to 100. Meanwhile, Firm Delta did not even exist in the first period, so the only way it can start production is to purchase permits to emit 50 tons of lead.

The total quantity of pollution will decline. But the buying and selling of the marketable permits will determine exactly which firms reduce pollution and by how much. With a system of marketable permits, the firms that find it least expensive to do so will reduce pollution the most.

	Firm Alpha	Firm Beta	Firm Gamma	Firm Delta
Current emissions—permits distributed free for this amount	200 tons	400 tons	600 tons	0 tons
How much pollution will these permits allow in one year?	100 tons	200 tons	300 tons	0 tons
Actual emissions one year in the future	150 tons	200 tons	200 tons	50 tons
Buyer or seller of marketable permit?	Buys permits for 50 tons	Doesn't buy or sell permits	Sells permits for 100 tons	Buys permits for 50 tons

How Marketable Permits Work

Another application of marketable permits occurred when the Clean Air Act was amended in 1990. The revised law sought to reduce sulfur dioxide emissions from electric power plants to half of the 1980 levels out of concern that sulfur dioxide was causing acid rain, which harms forests as well as buildings. In this case, the marketable permits the federal government issued were free of charge (no pun intended) to electricity-generating plants across the country, especially those that were burning coal (which produces sulfur dioxide). These permits were of the “shrinkable” type; that is, the amount of pollution allowed by a given permit declined with time.

Better-Defined Property Rights

A clarified and strengthened idea of property rights can also strike a balance between economic activity and pollution. Ronald Coase (1910–2013), who won the 1991 Nobel Prize in economics, offered a vivid illustration of an externality: a railroad track running beside a farmer’s field where the railroad locomotive sometimes gives off sparks and sets the field ablaze. Coase asked whose responsibility it was to address this spillover. Should the farmer be required to build a tall fence alongside the field to block the sparks? Or should the railroad be required to put some gadget on the locomotive’s smokestack to reduce the number of sparks?

Coase pointed out that this issue cannot be resolved until **property rights** are clearly defined—that is, the legal rights of ownership on which others are not allowed to infringe without paying compensation. Does the farmer have a property right not to have a field burned? Does the railroad have a property right to run its own trains on its own tracks? If neither party has a property right, then the two sides may squabble endlessly, nothing will be done, and sparks will continue to set the field aflame. However, if either the farmer or the railroad has a well-defined legal responsibility, then that party will seek out and pay for the least costly method of reducing the risk that sparks will hit the field. The property right determines whether the farmer or the railroad pays the bills.

The property rights approach is highly relevant in cases involving endangered species. The U.S. government’s endangered species list

includes about 1,000 plants and animals, and about 90% of these species live on privately owned land. The protection of these endangered species requires careful thinking about incentives and property rights. The discovery of an endangered species on private land has often triggered an automatic reaction from the government to prohibit the landowner from using that land for any purpose that might disturb the imperiled creatures. Consider the incentives of that policy: If you admit to the government that you have an endangered species, the government effectively prohibits you from using your land. As a result, rumors abounded of landowners who followed a policy of “shoot, shovel, and shut up” when they found an endangered animal on their land. Other landowners have deliberately cut trees or managed land in a way that they knew would discourage endangered animals from locating there.

Note:

How effective are market-oriented environmental policy tools?

Environmentalists sometimes fear that market-oriented environmental tools are an excuse to weaken or eliminate strict limits on pollution emissions and instead to allow more pollution. It is true that if pollution charges are set very low or if marketable permits do not reduce pollution by very much then market-oriented tools will not work well. But command-and-control environmental laws can also be full of loopholes or have exemptions that do not reduce pollution by much, either. The advantage of market-oriented environmental tools is not that they reduce pollution by more or less, but because of their incentives and flexibility, they can achieve any desired reduction in pollution at a lower cost to society.

A more productive policy would consider how to provide private landowners with an incentive to protect the endangered species that they find and to provide a habitat for additional endangered species. For example, the government might pay landowners who provide and maintain suitable habitats for endangered species or who restrict the use of their land to protect an endangered species. Again, an environmental law built on incentives and flexibility offers greater promise than a command-and-

control approach, which tries to oversee millions of acres of privately owned land.

Applying Market-Oriented Environmental Tools

Market-oriented environmental policies are a tool kit. Specific policy tools will work better in some situations than in others. For example, marketable permits work best when a few dozen or a few hundred parties are highly interested in trading, as in the cases of oil refineries that trade lead permits or electrical utilities that trade sulfur dioxide permits. However, for cases in which millions of users emit small amounts of pollution—such as emissions from car engines or unrecycled soda cans—and have no strong interest in trading, pollution charges will typically offer a better choice. Market-oriented environmental tools can also be combined. Marketable permits can be viewed as a form of improved property rights. Or the government could combine marketable permits with a pollution tax on any emissions not covered by a permit.

Key Concepts and Summary

Examples of market-oriented environmental policies, also referred to as cap and trade programs, include pollution charges, marketable permits, and better-defined property rights. Market-oriented environmental policies include taxes, markets, and property rights so that those who impose negative externalities must face the social cost.

Self-Check Questions

Exercise:

Problem:

Classify the following pollution-control policies as command-and-control or market incentive based.

- a. A state emissions tax on the quantity of carbon emitted by each firm.

- b. The federal government requires domestic auto companies to improve car emissions by 2020.
 - c. The EPA sets national standards for water quality.
 - d. A city sells permits to firms that allow them to emit a specified quantity of pollution.
 - e. The federal government pays fishermen to preserve salmon.
-

Solution:

- a. market-based
- b. command-and-control
- c. command-and-control
- d. market-based
- e. market-based

Exercise:

Problem:

An emissions tax on a quantity of emissions from a firm is not a command-and-control approach to reducing pollution. Why?

Solution:

Even though state or local governments impose these taxes, a company has the flexibility to adopt technologies that will help it avoid the tax.

Exercise:

Problem:

Four firms called Elm, Maple, Oak, and Cherry, produce wooden chairs. However, they also produce a great deal of garbage (a mixture of glue, varnish, sandpaper, and wood scraps). The first row of [\[link\]](#) shows the total amount of garbage (in tons) currently produced by each firm. The other rows of the table show the cost of reducing garbage produced by the first five tons, the second five tons, and so on. First, calculate the cost of requiring each firm to reduce the weight of its garbage by one-fourth. Now, imagine that marketable permits are issued for the current level of garbage, but the permits will shrink the weight of allowable garbage for each firm by one-fourth. What will be the result of this alternative approach to reducing pollution?

	Elm	Maple	Oak	Cherry
Current production of garbage (in tons)	20	40	60	80
Cost of reducing garbage by first five tons	\$5,500	\$6,300	\$7,200	\$3,000
Cost of reducing garbage by second five tons	\$6,000	\$7,200	\$7,500	\$4,000

	Elm	Maple	Oak	Cherry
Cost of reducing garbage by third five tons	\$6,500	\$8,100	\$7,800	\$5,000
Cost of reducing garbage by fourth five tons	\$7,000	\$9,000	\$8,100	\$6,000
Cost of reducing garbage by fifth five tons	\$0	\$9,900	\$8,400	\$7,000

Solution:

First, if each firm is required to reduce its garbage output by one-fourth, then Elm will reduce five tons at a cost of \$5,500; Maple will reduce 10 tons at a cost of \$13,500; Oak will reduce three tons at a cost of \$22,500; and Cherry will reduce four tons at a cost of \$18,000. Total cost of this approach: \$59,500. If the system of marketable permits is put in place, and those permits shrink the weight of allowable garbage by one-quarter, then pollution must still be reduced by the same overall amount. However, now the reduction in pollution will take place where it is least expensive.

Reductions in Garbage	Who does the reducing?	At what cost?
First 5 tons	Cherry	\$3,000
Second 5 tons	Cherry	\$4,000
Third 5 tons	Cherry	\$5,000
Fourth 5 tons	Elm	\$5,500
Fifth and sixth 5 tons	Elm and Cherry	\$6,000 each
Seventh 5 tons	Maple	\$6,300
Eighth 5 tons	Elm	\$6,500
Ninth and tenth 5 tons	Elm and Cherry	\$7,000 each

Thus, the overall pattern of reductions here will be that Elm reduces garbage by 20 tons and has 15 tons of permits to sell. Maple reduces by five tons and needs to buy five tons of permits. Oak does not reduce garbage at all, and needs to buy 15 tons of permits. Cherry reduces garbage by 25 tons, which leaves it with five tons of permits to sell. The total cost of these reductions would be \$56,300, a definite reduction in costs from the \$59,500 cost of the command-and-control option.

Exercise:

Problem:

The rows in [\[link\]](#) show three market-oriented tools for reducing pollution. The columns of the table show three complaints about command-and-control regulation. Fill in the table by stating briefly how each market-oriented tool addresses each of the three concerns.

	Incentives to Go Beyond	Flexibility about Where and How Pollution Will Be Reduced	Political Process Creates Loopholes and Exceptions
Pollution Charges			
Marketable Permits			
Property Rights			

Solution:

	Incentives to Go Beyond	Flexibility about Where and How Pollution Will Be Reduced	Political Process Creates Loopholes and Exceptions
Pollution Charges	If you keep reducing pollution you reduce your charge	Reducing pollution by any method is fine	If charge applies to all emissions of pollution then no loopholes
Marketable Permits	If you reduce your pollution you can sell your extra pollution permits	Reductions of pollution will happen at firms where it is cheapest to do so, by the least expensive methods	If all polluters are required to have permits then there are no loopholes
Property Rights	The party that has to pay for the pollution has incentive to do so in a cost effect way	Reducing pollution by any method is fine	If the property rights are clearly defined, then it is not legally possible to avoid cleanup

Review Questions

Exercise:

Problem:

What is a pollution charge and what incentive does it provide for a firm to take external costs into account?

Exercise:

Problem:

What is a marketable permit and what incentive does it provide for a firm to take external costs into account?

Exercise:

Problem:

What are better-defined property rights and what incentive do they provide to take external costs into account?

Critical Thinking Questions

Exercise:

Problem:

Will a system of marketable permits work with thousands of firms? Why or why not?

Exercise:

Problem:

Is zero pollution possible under a marketable permits system? Why or why not?

Exercise:

Problem: Is zero pollution an optimal goal? Why or why not?

References

Environmental Protection Agency. “2006 Pay-As-You-Throw Programs.” Accessed December 20, 2013.

<http://www.epa.gov/epawaste/conserve/tools/payt/states/06comm.htm>.

Glossary

marketable permit program

a permit that allows a firm to emit a certain amount of pollution; firms with more permits than pollution can sell the remaining permits to other firms

pollution charge

a tax imposed on the quantity of pollution that a firm emits; also called a pollution tax

property rights

the legal rights of ownership on which others are not allowed to infringe without paying compensation

The Benefits and Costs of U.S. Environmental Laws

By the end of this section, you will be able to:

- Evaluate the benefits and costs of environmental protection
- Explain the effects of ecotourism
- Apply marginal analysis to illustrate the marginal costs and marginal benefits of reducing pollution

Government economists have estimated that U.S. firms may pay more than \$200 billion per year to comply with federal environmental laws. That is big bucks. Is the money well spent?

Benefits and Costs of Clean Air and Clean Water

The benefits of a cleaner environment can be divided into four areas: (1) people may stay healthier and live longer; (2) certain industries that rely on clean air and water, such as farming, fishing, and tourism, may benefit; (3) property values may be higher; and (4) people may simply enjoy a cleaner environment in a way that does not need to involve a market transaction. Some of these benefits, such as gains to tourism or farming, are relatively easy to value in economic terms. It is harder to assign a monetary value to others, such as the value of clean air for someone with asthma. It seems impossible to put a clear-cut monetary value on still others, such as the satisfaction you might feel from knowing that the air is clear over the Grand Canyon, even if you have never visited the Grand Canyon.

Although estimates of environmental benefits are not precise, they can still be revealing. For example, a study by the Environmental Protection Agency looked at the costs and benefits of the Clean Air Act from 1970 to 1990. It found that total costs over that time period were roughly \$500 billion—a huge amount. However, it also found that a middle-range estimate of the health and other benefits from cleaner air was \$22 trillion—about 44 times higher than the costs. A more recent study by the EPA estimated that the environmental benefits to Americans from the Clean Air Act will exceed their costs by a margin of four to one. The EPA estimated that “in 2010 the benefits of Clean Air Act programs will total about \$110 billion. This

estimate represents the value of avoiding increases in illness and premature death which would have prevailed.” Saying that overall benefits of environmental regulation have exceeded costs in the past, however, is very different from saying that every environmental regulation makes sense. For example, studies suggest that when breaking down emission reductions by type of contaminants, the benefits of air pollution control outweigh the costs primarily for particulates and lead, but when looking at other air pollutants, the costs of reducing them may be comparable to or greater than the benefits.

Key Concepts and Summary

We can make a strong case, taken as a whole, that the benefits of U.S. environmental regulation have outweighed the costs. As the extent of environment regulation increases, additional expenditures on environmental protection will probably have increasing marginal costs and decreasing marginal benefits. This pattern suggests that the flexibility and cost savings of market-oriented environmental policies will become more important.

Self-Check Questions

Exercise:

Problem:

Suppose a city releases 16 million gallons of raw sewage into a nearby lake. [\[link\]](#) shows the total costs of cleaning up the sewage to different levels, together with the total benefits of doing so. (Benefits include environmental, recreational, health, and industrial benefits.)

	Total Cost (in thousands of dollars)	Total Benefits (in thousands of dollars)
16 million gallons	Current situation	Current situation
12 million gallons	50	800
8 million gallons	150	1300
4 million gallons	500	1650
0 gallons	1200	1900

- Using the information in [\[link\]](#), calculate the marginal costs and marginal benefits of reducing sewage emissions for this city. See [Cost and Industry Structure](#) if you need a refresher on how to calculate marginal costs.
- What is the optimal level of sewage for this city?
- Why not just pass a law that zero sewage can be emitted? After all, the total benefits of zero emissions exceed the total costs.

Solution:

- See the answers in the following table. The marginal cost is calculated as the change in total cost divided by the change in quantity.

	Total Cost (in thousands of dollars) [marginal cost]	Total Benefits (in thousands of dollars) [marginal benefit]
16 million gallons	Current situation	Current situation
12 million gallons	50 [50]	800 [800]
8 million gallons	150 [100]	1,300 [500]
4 million gallons	500 [350]	1,850 [350]
0 gallons	1,200 [700]	2,000 [150]

- b. The “optimal” level of pollution is where the marginal benefits of reducing it are equal to the marginal cost. This is at four million gallons.
- c. Marginal analysis tells us if the marginal costs of cleanup are greater than the marginal benefit, society could use those resources more efficiently elsewhere in the economy.

Exercise:

Problem:

The state of Colorado requires oil and gas companies who use fracking techniques to return the land to its original condition after the oil and gas extractions. [\[link\]](#) shows the total cost and total benefits (in dollars) of this policy.

Land Restored (in acres)	Total Cost	Total Benefit
0	\$0	\$0
100	\$20	\$140
200	\$80	\$240
300	\$160	\$320
400	\$280	\$380

- Calculate the marginal cost and the marginal benefit at each quantity (acre) of land restored. See [Cost and Industry Structure](#) if you need a refresher on how to calculate marginal costs and benefits.
- If we apply marginal analysis, what is the optimal amount of land to be restored?

Solution:

- See the next table for the answers, which were calculated using the traditional calculation of marginal cost equal to change in total

cost divided by change in quantity.

Land Restored (in acres)	Total Cost [marginal cost]	Total Benefit [marginal benefit]
0	\$0	\$0
100	\$20 [0.2]	\$140 [1.4]
200	\$80 [0.6]	\$240 [1]
300	\$160 [0.8]	\$320 [0.8]
400	\$280 [1.2]	\$480 [0.6]

- b. The optimal amount of restored land is 300 acres. Beyond this quantity the marginal costs are greater than the marginal benefits.

Review Questions

Exercise:

Problem:

As the extent of environmental protection expands, would you expect marginal costs of environmental protection to rise or fall? Why or why not?

Exercise:

Problem:

As the extent of environmental protection expands, would you expect the marginal benefits of environmental protection to rise or fall? Why or why not?

Critical Thinking Questions**Exercise:****Problem:**

From an economic perspective, is it sound policy to pursue a goal of zero pollution? Why or why not?

Exercise:**Problem:**

Recycling is a relatively inexpensive solution to much of the environmental contamination from plastics, glass, and other waste materials. Is it a sound policy to make it mandatory for everybody to recycle?

Problems**Exercise:****Problem:**

A city currently emits 16 million gallons (MG) of raw sewage into a lake that is beside the city. [\[link\]](#) shows the total costs (TC) in thousands of dollars of cleaning up the sewage to different levels, together with the total benefits (TB) of doing so. Benefits include environmental, recreational, health, and industrial benefits.

	TC	TB
16 MG	Current	Current
12 MG	50	800
8 MG	150	1300
4 MG	500	1850
0 MG	1200	2000

- Using the information in [\[link\]](#) calculate the marginal costs and marginal benefits of reducing sewage emissions for this city.
- What is the optimal level of sewage for this city? How can you tell?

References

Ryan, Dave. “New Report Shows Benefits of 1990 Clean Air Amendments Outweigh Costs by Four-to-One Margin,” press release, November 16, 1999. *United States Environmental Protection Agency*. Accessed December 19, 2013. <http://www.epa.gov/oar/sect812/r-140.html>.

National Center for Environmental Economics (NCEE). “Frequently Asked Questions on Mortality Risk Valuation.” *United States Environmental Protection Agency*. Accessed December 19, 2013.

<http://yosemite.epa.gov/ee/epa/eed.nsf/pages/MortalityRiskValuation.html#whatvalue> World Tourism Organization, “Tourism 2020 Vision.” Accessed December 19, 2013. http://www.world-tourism.org/market_research/facts/market_trends.htm.

Viscusi, Kip W. *Fatal Tradeoffs: Public and Private Responsibilities for Risk*. New York: Oxford University Press, 1995.

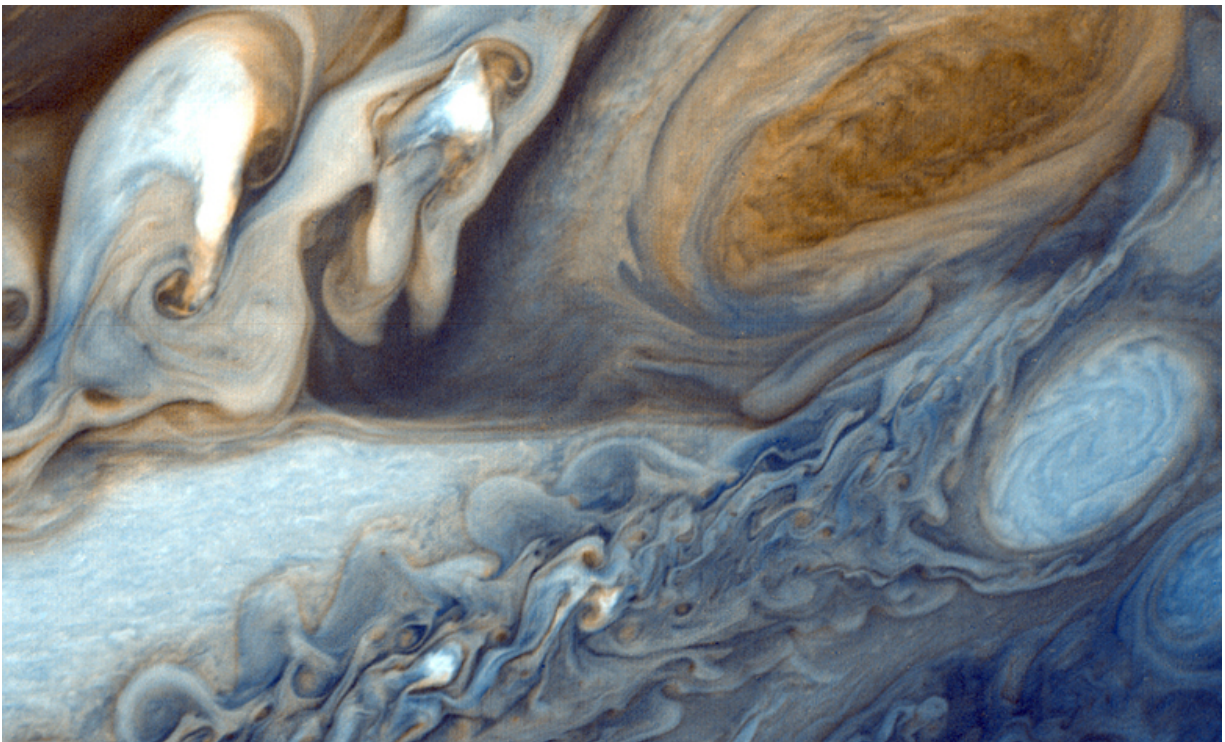
Introduction to Positive Externalities and Public Goods

class="introduction"

View from Voyager I

Launched by
NASA on
September 5,
1977,
Voyager 1's
primary
mission was
to provide
detailed
images of
Jupiter,
Saturn, and
their moons.
It took this
photograph
of Jupiter on
its journey.
In August of
2012,
Voyager I
entered
interstellar
space—the
first human-
made object
to do so—
and it is
expected to
send data
and images
back to earth
until 2025.
Such a

technological feat has a lot to do with economic principles.
(Credit: modification of work by NASA/JPL)



Note:

The Benefits of Voyager I Live On

The rapid growth of technology has increased our ability to access and process data, to navigate through a busy city, and to communicate with friends on the other side of the globe. The research and development efforts of citizens, scientists, firms, universities, and governments have

truly revolutionized the modern economy. To get a sense of how far we have come in a short period of time, let's compare one of humankind's greatest achievements to the smartphone most of us have in our coat pocket.

In 1977 the United States launched Voyager I, a spacecraft originally intended to reach Jupiter and Saturn, to send back photographs and other cosmic measurements. Voyager I, however, kept going, and going—past Jupiter and Saturn—right out of our solar system. At the time of its launch, Voyager had some of the most sophisticated computing processing power NASA could engineer (8,000 instructions per second), but by the time it left the solar system (in 2012, actually) we Earthlings were using handheld devices that could process 14 billion instructions per second.

Still, the technology of today is a spillover product of the incredible feats accomplished by NASA thirty years ago. NASA research, for instance, is responsible for the kidney dialysis and mammogram machines that we use today. Research in new technologies not only produces private benefits to the investing firm, or in this case to NASA, but it also creates benefits for the broader society. In this way, new knowledge often becomes what economists refer to as a public good. This leads us to the topic of this chapter—technology, positive externalities, public goods, and the role of government in the encouragement of innovation and the social benefits that it provides.

Note:**Introduction to Positive Externalities and Public Goods**

In this chapter, you will learn about:

- Why the Private Sector Under Invests in Technologies
- How Governments Can Encourage Innovation
- Public Goods

Can you imagine a world in which you did not own a cellular phone or use Wikipedia? New technology changes how people live and work and what

they buy. Technology includes the invention of new products, new ways of producing goods and services, and even new ways of managing a company more efficiently. Research and development of technology is the difference between horses and automobiles, between candles and electric lights, between fetching water in buckets and indoor plumbing, and between infection and good health from antibiotics.

In December 2009, ABC News compiled a list of some of the technological breakthroughs that have revolutionized consumer products in the past 10 years:

- GPS tracking devices, originally developed by the defense department and available to consumers in 2000, give users up-to-date information on location and time through satellite technology.
- In 2000, Toyota introduced the Prius hybrid car, which greatly improved fuel efficiency.
- Also in 2000, AT&T offered its customers the ability to text on a mobile phone.
- In 2001, Wikipedia launched a user-generated encyclopedia on the Web.
- Even though Napster died in 2001, the company launched music downloading and file sharing, which revolutionized how consumers get their music and videos.
- Friendster kicked off the social networking business in 2003, and Twitter and Facebook followed.
- In 2003, the Human Genome project was completed. It helps to fight disease and launch new pharmaceutical innovations.
- Also in 2003, the search engine became a way of life for obtaining information quickly. The search engine companies also became innovators in the digital software that dominates mobile devices.
- In 2006, Nintendo launched Wii and changed the way video games are played. Players can now be drawn into the action and use their bodies to respond rather than a handheld device.
- Apple introduced the iPhone in 2007 and launched an entire smartphone industry. In 2015, cell phones now recognize human voices via artificial intelligence.

With all new technologies, however, there are new challenges. This chapter deals with some of these issues: Will private companies be willing to invest in new technology? In what ways does new technology have positive externalities? What motivates inventors? Does government have a role to play in encouraging research and technology? Are there certain types of goods that markets fail to provide efficiently, and that only government can produce? What happens when consumption or production of a product creates positive externalities? Why is it unsurprising when a common resource, like marine fisheries, is overused?

Why the Private Sector Under Invests in Innovation

By the end of this section, you will be able to:

- Identify the positive externalities of new technology.
- Explain the difference between private benefits and social benefits and give examples of each.
- Calculate and analyze rates of return

Market competition can provide an incentive for discovering new technology because a firm can earn higher profits by finding a way to produce products more cheaply or to create products with characteristics consumers want. As Gregory Lee, CEO of Samsung said, “Relentless pursuit of new innovation is the key principle of our business and enables consumers to discover a world of possibilities with technology.” An innovative firm knows that it will usually have a temporary edge over its competitors and thus an ability to earn above-normal profits before competitors can catch up.

In certain cases, however, competition can discourage new technology, especially when other firms can quickly copy a new idea. Consider a pharmaceutical firm deciding to develop a new drug. On average, it can cost \$800 million and take more than a decade to discover a new drug, perform the necessary safety tests, and bring the drug to market. If the research and development (R&D) effort fails—and every R&D project has some chance of failure—then the firm will suffer losses and could even be driven out of business. If the project succeeds, then the firm’s competitors may figure out ways of adapting and copying the underlying idea, but without having to pay the costs themselves. As a result, the innovative company will bear the much higher costs of the R&D and will enjoy at best only a small, temporary advantage over the competition.

Many inventors over the years have discovered that their inventions brought them less profit than they might have reasonably expected.

- Eli Whitney (1765–1825) invented the cotton gin, but then southern cotton planters built their own seed-separating devices with a few minor changes in Whitney’s design. When Whitney sued, he found that the courts in southern states would not uphold his patent rights.

- Thomas Edison (1847–1931) still holds the record for most patents granted to an individual. His first invention was an automatic vote counter, and despite the social benefits, he could not find a government that wanted to buy it.
- Gordon Gould came up with the idea behind the laser in 1957. He put off applying for a patent and, by the time he did apply, other scientists had laser inventions of their own. A lengthy legal battle resulted, in which Gould spent \$100,000 on lawyers, before he eventually received a patent for the laser in 1977. Compared to the enormous social benefits of the laser, Gould received relatively little financial reward.
- In 1936, Turing delivered a paper titled, "On Computable Numbers, with an Application to the Entscheidungsproblem," in which he presented the notion of a universal machine (later called the "Universal Turing Machine," and then the "Turing machine") capable of computing anything that is computable. The central concept of the modern computer was based on Turing's paper.

A variety of studies by economists have found that the original inventor receives one-third to one-half of the total economic benefits from innovations, while other businesses and new product users receive the rest.

The Positive Externalities of New Technology

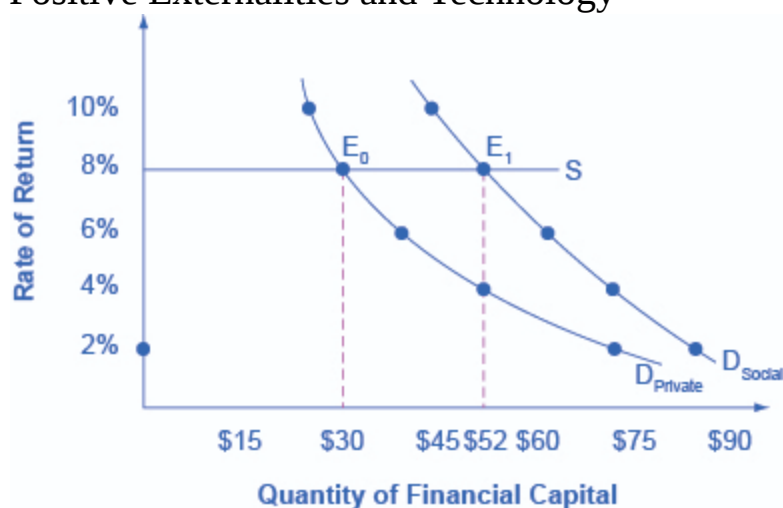
Will private firms in a market economy under invest in research and technology? If a firm builds a factory or buys a piece of equipment, the firm receives all the economic benefits that result from the investments. However, when a firm invests in new technology, the **private benefits**, or profits, that the firm receives are only a portion of the overall social benefits. The **social benefits** of an innovation take into account the value of all the positive externalities of the new idea or product, whether enjoyed by other companies or society as a whole, as well as the private benefits received by the firm that developed the new technology. As you learned in [Environmental Protection and Negative Externalities](#), **positive externalities** are beneficial spillovers to a third party, or parties.

Consider the example of the Big Drug Company, which is planning its R&D budget for the next year. Economists and scientists working for Big

Drug have compiled a list of potential research and development projects and estimated rates of return. (The rate of return is the estimated payoff from the project.) [\[link\]](#) shows how the calculations work. The downward-sloping D_{Private} curve represents the firm's demand for financial capital and reflects the company's willingness to borrow to finance research and development projects at various interest rates. Suppose that this firm's investment in research and development creates a spillover benefit to other firms and households. After all, new innovations often spark other creative endeavors that society also values. If we add the spillover benefits society enjoys to the firm's private demand for financial capital, we can draw D_{Social} that lies above D_{Private} .

If there was a way for the firm to fully monopolize those social benefits by somehow making them unavailable to the rest of us, the firm's private demand curve would be the same as society's demand curve. According to [\[link\]](#) and [\[link\]](#), if the going rate of interest on borrowing is 8%, and the company can receive the private benefits of innovation only, then the company would finance \$30 million. Society, at the same rate of 8%, would find it optimal to have \$52 million of borrowing. Unless there is a way for the company to fully enjoy the total benefits, then it will borrow less than the socially optimal level of \$52 million.

Positive Externalities and Technology



Big Drug faces a cost of borrowing of 8%.

If the firm receives only the private benefits of investing in R&D, then its

demand curve for financial capital is shown by D_{Private} , and the equilibrium will occur at \$30 million. Because there are spillover benefits, society would find it optimal to have \$52 million of investment. If the firm could keep the social benefits of its investment for itself, its demand curve for financial capital would be D_{Social} and it would be willing to borrow \$52 million.

Rate of Return	D_{Private} (in millions)	D_{Social} (in millions)
2%	\$72	\$84
4%	\$52	\$72
6%	\$38	\$62
8%	\$30	\$52
10%	\$26	\$44

Return and Demand for Capital

Big Drug's original demand for financial capital (D_{Private}) is based on the profits received by the firm. However, other pharmaceutical firms and health care companies may learn new lessons about how to treat certain medical conditions and are then able to create their own competing products. The social benefit of the drug takes into account the value of all the positive externalities of the drug. If Big Drug were able to gain this

social return instead of other companies, its demand for financial capital would shift to the demand curve D_{Social} , and it would be willing to borrow and invest \$52 million. However, if Big Drug is receiving only 50 cents of each dollar of social benefits, the firm will not spend as much on creating new products. The amount it would be willing to spend would fall somewhere in between D_{Private} and D_{Social} .

Why Invest in Human Capital?

The investment in anything, whether it is the construction of a new power plant or research in a new cancer treatment, usually requires a certain upfront cost with an uncertain future benefit. The investment in education, or human capital, is no different. Over the span of many years, a student and her family invest significant amounts of time and money into that student's education. The idea is that higher levels of educational attainment will eventually serve to increase the student's future productivity and subsequent ability to earn. Once the numbers are crunched, does this investment pay off for the student?

Almost universally, economists have found that the answer to this question is a clear "Yes." For example, several studies of the return to education in the United States estimate that the rate of return to a college education is approximately 10%. Data in [\[link\]](#), from the U.S. Bureau of Labor Statistics' *Usual Weekly Earnings of Wage and Salary Workers, Third Quarter 2014*, demonstrate that median weekly earnings are higher for workers who have completed more education. While these rates of return will beat equivalent investments in Treasury bonds or savings accounts, the estimated returns to education go primarily to the individual worker, so these returns are **private rates of return** to education.

	Less than a High School Degree	High School Degree, No College	Bachelor's Degree
Median Weekly Earnings (full-time workers over the age of 25)	\$488	\$668	\$1,101

Usual Weekly Earnings of Wage and Salary Workers, Third Quarter 2014(Source: <http://www.bls.gov/news.release/pdf/wkyeng.pdf>)

What does society gain from investing in the education of another student? After all, if the government is spending taxpayer dollars to subsidize public education, society should expect some kind of return on that spending. Again, economists like George Psacharopoulos have found that, across a variety of nations, the **social rate of return** on schooling is also positive. After all, positive externalities exist from investment in education. While not always easy to measure, according to Walter McMahon, the positive externalities to education typically include better health outcomes for the population, lower levels of crime, a cleaner environment and a more stable, democratic government. For these reasons, many nations have chosen to use taxpayer dollars to subsidize primary, secondary, and higher education. Education clearly benefits the person who receives it, but a society where most people have a good level of education provides positive externalities for all.

Other Examples of Positive Externalities

Although technology may be the most prominent example of a positive externality, it is not the only one. For example, being vaccinated against disease is not only a protection for the individual, but it has the positive spillover of protecting others who may become infected. When a number of homes in a neighborhood are modernized, updated, and restored, not only

does it increase the value of those homes, but the value of other properties in the neighborhood may increase as well.

The appropriate public policy response to a positive externality, like a new technology, is to help the party creating the positive externality receive a greater share of the social benefits. In the case of vaccines, like flu shots, an effective policy might be to provide a subsidy to those who choose to get vaccinated.

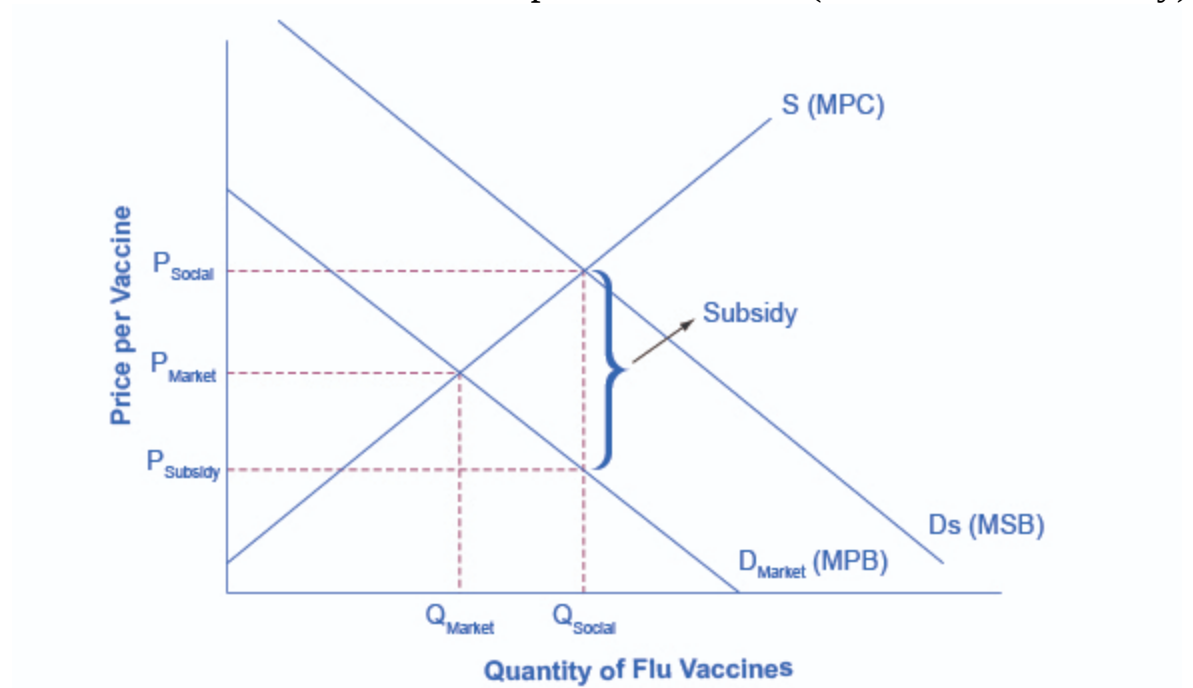
[\[link\]](#) shows the market for flu shots. The market demand curve D_{Market} for flu shots reflects only the marginal private benefits (MPB) that the vaccinated individuals receive from the shots. Assuming that there are no spillover costs in the production of flu shots, the market supply curve is given by the marginal private cost (MPC) of producing the vaccinations.

The equilibrium quantity of flu shots produced in the market, where MPB is equal to MPC, is Q_{Market} and the price of flu shots is P_{Market} . However, spillover benefits exist in this market because others, those who chose not to purchase a flu shot, receive a positive externality in a reduced chance of contracting the flu. When we add the spillover benefits to the marginal private benefit of flu shots, the marginal social benefit (MSB) of flu shots is given by D_{Social} . Because the MPB is greater than MSB, we see that the socially optimal level of flu shots is greater than the market quantity (Q_{Social} exceeds Q_{Market}) and the corresponding price of flu shots, if the market were to produce Q_{Social} , would be at P_{Social} . Unfortunately, the marketplace does not recognize the positive externality and flu shots will go under produced and under consumed.

So how can government try to move the market level of output closer to the socially desirable level of output? One policy would be to provide a subsidy, like a voucher, to any citizen who wishes to get vaccinated. This voucher would act as “income” that could be used to purchase only a flu shot and, if the voucher was exactly equal to the per-unit spillover benefits, would increase market equilibrium to a quantity of Q_{Social} and a price of P_{Social} where MSB equals MSC. Suppliers of the flu shots would receive payment of P_{Social} per vaccination, while consumers of flu shots would redeem the voucher and only pay a price of P_{Subsidy} . When the government

uses a subsidy in this way, the socially optimal quantity of vaccinations is produced.

The Market for Flu Shots with Spillover Benefits (A Positive Externality)



The market demand curve does not reflect the positive externality of flu vaccinations, so only Q_{Market} will be exchanged. This outcome is inefficient because the marginal social benefit exceeds the marginal social cost. If the government provides a subsidy to consumers of flu shots, equal to the marginal social benefit minus the marginal private benefit, the level of vaccinations can increase to the socially optimal quantity of Q_{Social} .

Key Concepts and Summary

Competition creates pressure to innovate. However, if new inventions can be easily copied, then the original inventor loses the incentive to invest further in research and development. New technology often has positive externalities; that is, there are often spillovers from the invention of new technology that benefit firms other than the innovator. The social benefit of

an invention, once these spillovers are taken into account, typically exceeds the private benefit to the inventor. If inventors could receive a greater share of the broader social benefits for their work, they would have a greater incentive to seek out new inventions.

Self-Check Questions

Exercise:

Problem:

Are positive externalities reflected in market demand curves? Why or why not?

Solution:

No. A market demand curve reflects only the private benefits of those who are consuming the product. Positive externalities are benefits that spill over to third parties, so they create social benefits, and are not captured by a market (or private benefit) demand curve.

Exercise:

Problem:

Samsung's R&D investment in digital devices has increased profits by 20%. Is this a private or social benefit?

Solution:

Clearly Samsung is benefiting from the investment, so the 20% increase in profits is a private benefit. If Samsung is unable to capture all of the benefit, perhaps because other companies quickly copy and produce close substitutes, then Samsung's investment will produce social benefits.

Exercise:

Problem:

The Gizmo Company is planning to develop new household gadgets. [\[link\]](#) shows the company's demand for financial capital for research and development of these gadgets, based on expected rates of return from sales. Now, say that every investment would have an additional 5% social benefit—that is, an investment that pays at least a 6% return to the Gizmo Company will pay at least an 11% return for society as a whole; an investment that pays at least 7% for the Gizmo Company will pay at least 12% for society as a whole, and so on. Answer the questions that follow based on this information.

Estimated Rate of Return	Private profits of the firm from an R&D project (in \$ millions)
10%	\$100
9%	\$102
8%	\$108
7%	\$118
6%	\$133
5%	\$153
4%	\$183
3%	\$223

- a. If the going interest rate is 9%, how much will Gizmo invest in R&D if it receives only the private benefits of this investment?
 - b. Assume that the interest rate is still 9%. How much will the firm invest if it also receives the social benefits of its investment? (Add an additional 5% return on all levels of investment.)
-

Solution:

- a. \$102 million.
- b. If the interest rate is 9%, the cost of financial capital, and the firm can capture the 5% return to society, the firm would invest as if its effective rate of return is 4%, so it will invest \$183 million.

Exercise:

Problem:

The Junkbuyers Company travels from home to home, looking for opportunities to buy items that would otherwise be put out with the garbage, but which the company can resell or recycle. Which will be larger, the private or the social benefits?

Solution:

When the Junkbuyers Company purchases something for resale, presumably both the buyer and the seller benefit—otherwise, they would not need to make the transaction. However, the company also reduces the amount of garbage produced, which saves money for households and/or for the city that disposes of garbage. So the social benefits are larger than the private benefits.

Review Questions

Exercise:

Problem:

In what ways do company investments in research and development create positive externalities?

Exercise:**Problem:**

Will the demand for borrowing and investing in R&D be higher or lower if there are no external benefits?

Critical Thinking Question**Exercise:****Problem:**

Can a company be guaranteed all of the social benefits of a new invention? Why or why not?

Problems**Exercise:**

Problem:

HighFlyer Airlines wants to build new airplanes with greatly increased cabin space. This will allow HighFlyer Airlines to give passengers more comfort and sell more tickets at a higher price. However, redesigning the cabin means rethinking many other elements of the airplane as well, like the placement of engines and luggage, and the most efficient shape of the plane for moving through the air. HighFlyer Airlines has developed a list of possible methods to increase cabin space, along with estimates of how these approaches would affect costs of operating the plane and sales of airline tickets. Based on these estimates, [\[link\]](#) shows the value of R&D projects that provide at least a certain private rate of return. Column 1 = Private Rate of Return. Column 2 = Value of R&D Projects that Return at Least the Private Rate of Return to HighFlyer Airlines. Use the data to answer the following questions.

Private Rate of Return	Value of R&D
12%	\$100
10%	\$200
8%	\$300
6%	\$400
4%	\$500

- a. If the opportunity cost of financial capital for HighFlyer Airlines is 6%, how much should the firm invest in R&D?

- b. Assume that the social rate of return for R&D is an additional 2% on top of the private return; that is, an R&D investment that had a 7% private return to HighFlyer Airlines would have a 9% social return. How much investment is socially optimal at the 6% interest rate?

References

Arias, Omar and Walter W. McMahon. “Dynamic Rates of Return to Education in the U.S.” *Economics of Education Review*. 20, 2001. 121–138.

Biography.com. 2015. “Alan Turing.” Accessed April 1, 2015.
<http://www.biography.com/people/alan-turing-9512017>.

Canty Media. 2015. “The World: Life Expectancy (2015) - Top 100+.” Accessed April 1, 2015. <http://www.geoba.se/population.php?pc=world&type=15>.

Hyclak, Thomas, Geraint Johnes, and Robert Thornton. *Fundamentals of Labor Economics*. Boston: Houghton Mifflin Company, 2005.

McMahon, Walter. *Education and Development: Measuring the Social Benefits*. Oxford: Oxford University Press, 2000.

National Institute of Health. 2015. “Global Competitiveness—The Importance of U.S. Leadership in Science and Innovation for the Future of Our Economy and Our Health.” Accessed April 1, 2015.
http://www.nih.gov/about/impact/impact_global.pdf.

National Science Foundation. 2013. “U.S. R&D Spending Resumes Growth in 2010 and 2011 but Still Lags Behind the Pace of Expansion of the National Economy.” Accessed April 1, 2015.
<http://www.nsf.gov/statistics/infbrief/nsf13313/>.

Psacharopoulos, George. “Returns to Investment in Education: A Global Update.” *World Development* 22, 1994. 1325–1343.

Salientes-Narisma, Corrie. "Samsung Shift to Innovative Devices Pay Off." *Inquirer Technology*. Accessed May 15, 2013.
<http://technology.inquirer.net/23831/samsungs-shift-to-innovative-devices-pays-off>.

Glossary

positive externalities

beneficial spillovers to a third party or parties

private benefits

the dollar value of all benefits of a new product or process invented by a company that can be captured by the investing company

private rates of return

when the estimated rates of return go primarily to an individual; for example, earning interest on a savings account

social benefits

the dollar value of all benefits of a new product or process invented by a company that can be captured by other firms and by society as a whole

social rate of return

when the estimated rates of return go primarily to society; for example, providing free education

Why the Private Sector Under Invests in Innovation

By the end of this section, you will be able to:

- Identify the positive externalities of new technology.
- Explain the difference between private benefits and social benefits and give examples of each.
- Calculate and analyze rates of return

Market competition can provide an incentive for discovering new technology because a firm can earn higher profits by finding a way to produce products more cheaply or to create products with characteristics consumers want. As Gregory Lee, CEO of Samsung said, “Relentless pursuit of new innovation is the key principle of our business and enables consumers to discover a world of possibilities with technology.” An innovative firm knows that it will usually have a temporary edge over its competitors and thus an ability to earn above-normal profits before competitors can catch up.

In certain cases, however, competition can discourage new technology, especially when other firms can quickly copy a new idea. Consider a pharmaceutical firm deciding to develop a new drug. On average, it can cost \$800 million and take more than a decade to discover a new drug, perform the necessary safety tests, and bring the drug to market. If the research and development (R&D) effort fails—and every R&D project has some chance of failure—then the firm will suffer losses and could even be driven out of business. If the project succeeds, then the firm’s competitors may figure out ways of adapting and copying the underlying idea, but without having to pay the costs themselves. As a result, the innovative company will bear the much higher costs of the R&D and will enjoy at best only a small, temporary advantage over the competition.

Many inventors over the years have discovered that their inventions brought them less profit than they might have reasonably expected.

- Eli Whitney (1765–1825) invented the cotton gin, but then southern cotton planters built their own seed-separating devices with a few minor changes in Whitney’s design. When Whitney sued, he found that the courts in southern states would not uphold his patent rights.

- Thomas Edison (1847–1931) still holds the record for most patents granted to an individual. His first invention was an automatic vote counter, and despite the social benefits, he could not find a government that wanted to buy it.
- Gordon Gould came up with the idea behind the laser in 1957. He put off applying for a patent and, by the time he did apply, other scientists had laser inventions of their own. A lengthy legal battle resulted, in which Gould spent \$100,000 on lawyers, before he eventually received a patent for the laser in 1977. Compared to the enormous social benefits of the laser, Gould received relatively little financial reward.
- In 1936, Turing delivered a paper titled, "On Computable Numbers, with an Application to the Entscheidungsproblem," in which he presented the notion of a universal machine (later called the "Universal Turing Machine," and then the "Turing machine") capable of computing anything that is computable. The central concept of the modern computer was based on Turing's paper.

A variety of studies by economists have found that the original inventor receives one-third to one-half of the total economic benefits from innovations, while other businesses and new product users receive the rest.

The Positive Externalities of New Technology

Will private firms in a market economy under invest in research and technology? If a firm builds a factory or buys a piece of equipment, the firm receives all the economic benefits that result from the investments. However, when a firm invests in new technology, the **private benefits**, or profits, that the firm receives are only a portion of the overall social benefits. The **social benefits** of an innovation take into account the value of all the positive externalities of the new idea or product, whether enjoyed by other companies or society as a whole, as well as the private benefits received by the firm that developed the new technology. As you learned in [Environmental Protection and Negative Externalities](#), **positive externalities** are beneficial spillovers to a third party, or parties.

Why Invest in Human Capital?

The investment in anything, whether it is the construction of a new power plant or research in a new cancer treatment, usually requires a certain upfront cost with an uncertain future benefit. The investment in education, or human capital, is no different. Over the span of many years, a student and her family invest significant amounts of time and money into that student’s education. The idea is that higher levels of educational attainment will eventually serve to increase the student’s future productivity and subsequent ability to earn. Once the numbers are crunched, does this investment pay off for the student?

Almost universally, economists have found that the answer to this question is a clear “Yes.” For example, several studies of the return to education in the United States estimate that the rate of return to a college education is approximately 10%. Data in [\[link\]](#), from the U.S. Bureau of Labor Statistics’ *Usual Weekly Earnings of Wage and Salary Workers, Third Quarter 2014*, demonstrate that median weekly earnings are higher for workers who have completed more education. While these rates of return will beat equivalent investments in Treasury bonds or savings accounts, the estimated returns to education go primarily to the individual worker, so these returns are **private rates of return** to education.

	Less than a High School Degree	High School Degree, No College	Bachelor’s Degree
Median Weekly Earnings (full-time workers over the age of 25)	\$488	\$668	\$1,101

Usual Weekly Earnings of Wage and Salary Workers, Third Quarter

2014(Source: <http://www.bls.gov/news.release/pdf/wkyeng.pdf>)

What does society gain from investing in the education of another student? After all, if the government is spending taxpayer dollars to subsidize public education, society should expect some kind of return on that spending. Again, economists like George Psacharopoulos have found that, across a variety of nations, the **social rate of return** on schooling is also positive. After all, positive externalities exist from investment in education. While not always easy to measure, according to Walter McMahon, the positive externalities to education typically include better health outcomes for the population, lower levels of crime, a cleaner environment and a more stable, democratic government. For these reasons, many nations have chosen to use taxpayer dollars to subsidize primary, secondary, and higher education. Education clearly benefits the person who receives it, but a society where most people have a good level of education provides positive externalities for all.

Other Examples of Positive Externalities

Although technology may be the most prominent example of a positive externality, it is not the only one. For example, being vaccinated against disease is not only a protection for the individual, but it has the positive spillover of protecting others who may become infected. When a number of homes in a neighborhood are modernized, updated, and restored, not only does it increase the value of those homes, but the value of other properties in the neighborhood may increase as well.

The appropriate public policy response to a positive externality, like a new technology, is to help the party creating the positive externality receive a greater share of the social benefits. In the case of vaccines, like flu shots, an effective policy might be to provide a subsidy to those who choose to get vaccinated.

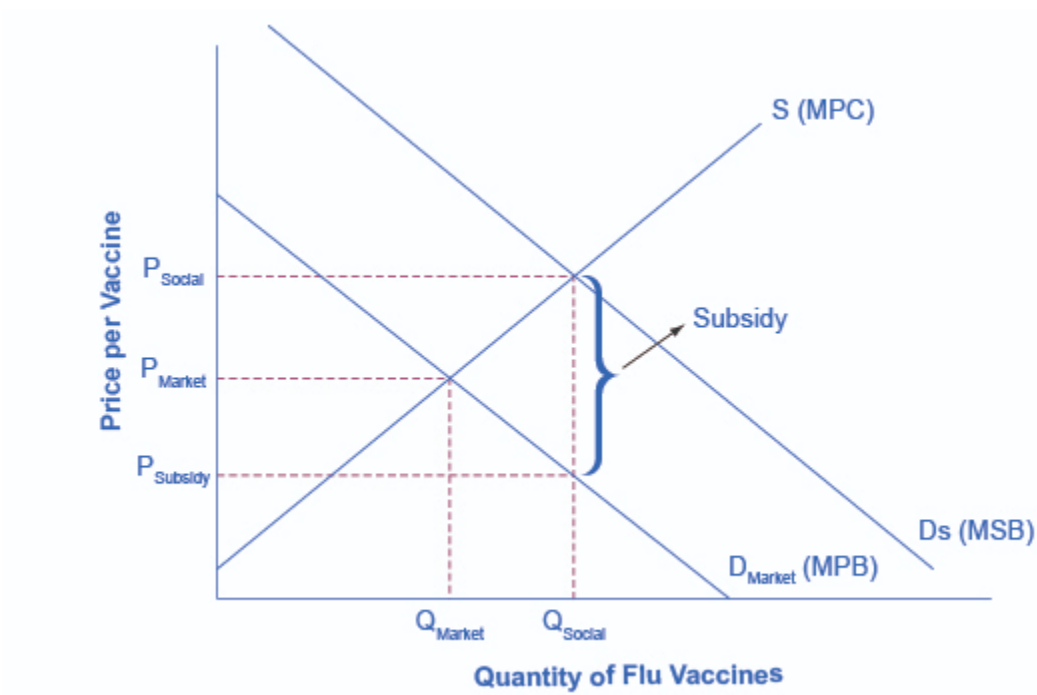
[\[link\]](#) shows the market for flu shots. The market demand curve D_{Market} for flu shots reflects only the marginal private benefits (MPB) that the vaccinated individuals receive from the shots. Assuming that there are no

spillover costs in the production of flu shots, the market supply curve is given by the marginal private cost (MPC) of producing the vaccinations.

The equilibrium quantity of flu shots produced in the market, where MPB is equal to MPC, is Q_{Market} and the price of flu shots is P_{Market} . However, spillover benefits exist in this market because others, those who chose not to purchase a flu shot, receive a positive externality in a reduced chance of contracting the flu. When we add the spillover benefits to the marginal private benefit of flu shots, the marginal social benefit (MSB) of flu shots is given by D_{Social} . Because the MPB is greater than MSB, we see that the socially optimal level of flu shots is greater than the market quantity (Q_{Social} exceeds Q_{Market}) and the corresponding price of flu shots, if the market were to produce Q_{Social} , would be at P_{Social} . Unfortunately, the marketplace does not recognize the positive externality and flu shots will go under produced and under consumed.

So how can government try to move the market level of output closer to the socially desirable level of output? One policy would be to provide a subsidy, like a voucher, to any citizen who wishes to get vaccinated. This voucher would act as “income” that could be used to purchase only a flu shot and, if the voucher was exactly equal to the per-unit spillover benefits, would increase market equilibrium to a quantity of Q_{Social} and a price of P_{Social} where MSB equals MSC. Suppliers of the flu shots would receive payment of P_{Social} per vaccination, while consumers of flu shots would redeem the voucher and only pay a price of P_{Subsidy} . When the government uses a subsidy in this way, the socially optimal quantity of vaccinations is produced.

The Market for Flu Shots with Spillover Benefits (A Positive Externality)



The market demand curve does not reflect the positive externality of flu vaccinations, so only Q_{Market} will be exchanged. This outcome is inefficient because the marginal social benefit exceeds the marginal social cost. If the government provides a subsidy to consumers of flu shots, equal to the marginal social benefit minus the marginal private benefit, the level of vaccinations can increase to the socially optimal quantity of Q_{Social} .

Key Concepts and Summary

Competition creates pressure to innovate. However, if new inventions can be easily copied, then the original inventor loses the incentive to invest further in research and development. New technology often has positive externalities; that is, there are often spillovers from the invention of new technology that benefit firms other than the innovator. The social benefit of an invention, once these spillovers are taken into account, typically exceeds the private benefit to the inventor. If inventors could receive a greater share

of the broader social benefits for their work, they would have a greater incentive to seek out new inventions.

Self-Check Questions

Exercise:

Problem:

Are positive externalities reflected in market demand curves? Why or why not?

Solution:

No. A market demand curve reflects only the private benefits of those who are consuming the product. Positive externalities are benefits that spill over to third parties, so they create social benefits, and are not captured by a market (or private benefit) demand curve.

Exercise:

Problem:

Samsung's R&D investment in digital devices has increased profits by 20%. Is this a private or social benefit?

Solution:

Clearly Samsung is benefiting from the investment, so the 20% increase in profits is a private benefit. If Samsung is unable to capture all of the benefit, perhaps because other companies quickly copy and produce close substitutes, then Samsung's investment will produce social benefits.

Exercise:

Problem:

The Gizmo Company is planning to develop new household gadgets. [\[link\]](#) shows the company's demand for financial capital for research and development of these gadgets, based on expected rates of return from sales. Now, say that every investment would have an additional 5% social benefit—that is, an investment that pays at least a 6% return to the Gizmo Company will pay at least an 11% return for society as a whole; an investment that pays at least 7% for the Gizmo Company will pay at least 12% for society as a whole, and so on. Answer the questions that follow based on this information.

Estimated Rate of Return	Private profits of the firm from an R&D project (in \$ millions)
10%	\$100
9%	\$102
8%	\$108
7%	\$118
6%	\$133
5%	\$153
4%	\$183
3%	\$223

- a. If the going interest rate is 9%, how much will Gizmo invest in R&D if it receives only the private benefits of this investment?
 - b. Assume that the interest rate is still 9%. How much will the firm invest if it also receives the social benefits of its investment? (Add an additional 5% return on all levels of investment.)
-

Solution:

- a. \$102 million.
- b. If the interest rate is 9%, the cost of financial capital, and the firm can capture the 5% return to society, the firm would invest as if its effective rate of return is 4%, so it will invest \$183 million.

Exercise:

Problem:

The Junkbuyers Company travels from home to home, looking for opportunities to buy items that would otherwise be put out with the garbage, but which the company can resell or recycle. Which will be larger, the private or the social benefits?

Solution:

When the Junkbuyers Company purchases something for resale, presumably both the buyer and the seller benefit—otherwise, they would not need to make the transaction. However, the company also reduces the amount of garbage produced, which saves money for households and/or for the city that disposes of garbage. So the social benefits are larger than the private benefits.

Review Questions

Exercise:

Problem:

In what ways do company investments in research and development create positive externalities?

Exercise:**Problem:**

Will the demand for borrowing and investing in R&D be higher or lower if there are no external benefits?

Critical Thinking Question**Exercise:****Problem:**

Can a company be guaranteed all of the social benefits of a new invention? Why or why not?

Problems**Exercise:**

Problem:

HighFlyer Airlines wants to build new airplanes with greatly increased cabin space. This will allow HighFlyer Airlines to give passengers more comfort and sell more tickets at a higher price. However, redesigning the cabin means rethinking many other elements of the airplane as well, like the placement of engines and luggage, and the most efficient shape of the plane for moving through the air. HighFlyer Airlines has developed a list of possible methods to increase cabin space, along with estimates of how these approaches would affect costs of operating the plane and sales of airline tickets. Based on these estimates, [\[link\]](#) shows the value of R&D projects that provide at least a certain private rate of return. Column 1 = Private Rate of Return. Column 2 = Value of R&D Projects that Return at Least the Private Rate of Return to HighFlyer Airlines. Use the data to answer the following questions.

Private Rate of Return	Value of R&D
12%	\$100
10%	\$200
8%	\$300
6%	\$400
4%	\$500

- a. If the opportunity cost of financial capital for HighFlyer Airlines is 6%, how much should the firm invest in R&D?

- b. Assume that the social rate of return for R&D is an additional 2% on top of the private return; that is, an R&D investment that had a 7% private return to HighFlyer Airlines would have a 9% social return. How much investment is socially optimal at the 6% interest rate?

References

Arias, Omar and Walter W. McMahon. "Dynamic Rates of Return to Education in the U.S." *Economics of Education Review*. 20, 2001. 121–138.

Biography.com. 2015. "Alan Turing." Accessed April 1, 2015. <http://www.biography.com/people/alan-turing-9512017>.

Canty Media. 2015. "The World: Life Expectancy (2015) - Top 100+." Accessed April 1, 2015. <http://www.geoba.se/population.php?pc=world&type=15>.

Hyclak, Thomas, Geraint Johnes, and Robert Thornton. *Fundamentals of Labor Economics*. Boston: Houghton Mifflin Company, 2005.

McMahon, Walter. *Education and Development: Measuring the Social Benefits*. Oxford: Oxford University Press, 2000.

National Institute of Health. 2015. "Global Competitiveness—The Importance of U.S. Leadership in Science and Innovation for the Future of Our Economy and Our Health." Accessed April 1, 2015. http://www.nih.gov/about/impact/impact_global.pdf.

National Science Foundation. 2013. "U.S. R&D Spending Resumes Growth in 2010 and 2011 but Still Lags Behind the Pace of Expansion of the National Economy." Accessed April 1, 2015. <http://www.nsf.gov/statistics/infbrief/nsf13313/>.

Psacharopoulos, George. "Returns to Investment in Education: A Global Update." *World Development* 22, 1994. 1325–1343.

Salientes-Narisma, Corrie. "Samsung Shift to Innovative Devices Pay Off." *Inquirer Technology*. Accessed May 15, 2013.
<http://technology.inquirer.net/23831/samsungs-shift-to-innovative-devices-pays-off>.

Glossary

positive externalities

beneficial spillovers to a third party or parties

private benefits

the dollar value of all benefits of a new product or process invented by a company that can be captured by the investing company

private rates of return

when the estimated rates of return go primarily to an individual; for example, earning interest on a savings account

social benefits

the dollar value of all benefits of a new product or process invented by a company that can be captured by other firms and by society as a whole

social rate of return

when the estimated rates of return go primarily to society; for example, providing free education

Introduction to Information, Risk, and Insurance

class="introduction"

President Obama's Health Care Reform

The Patient
Protection and
Affordable
Care Act has
become a
controversial
topic—one
which relates
strongly to the
topic of this
chapter.

(Credit:
modification
of work by
Daniel
Borman/Flick
r Creative
Commons)

**Note:****What's the Big Deal with Obamacare?**

In August 2009, many members of the U.S. Congress used their summer recess to return to their home districts and hold town hall-style meetings to discuss President Obama's proposed changes to the U.S. healthcare system. This was officially known as the Patient Protection and Affordable Care Act (PPACA) or as the Affordable Care Act (ACA), but was more popularly known as Obamacare. The bill's opponents' claims ranged from the charge that the changes were unconstitutional and would add \$750 billion to the deficit, to extreme claims about the inclusion of things like the implantation of microchips and so-called "death panels" that decide which critically-ill patients receive care and which do not.

Why did people react so strongly? After all, the intent of the law is to make healthcare insurance more affordable, to allow more people to get insurance, and to reduce the costs of healthcare. For each year from 2000 to 2011, these costs grew at least double the rate of inflation. In 2014, healthcare spending accounted for around 24% of all federal government spending. In the United States, we spend more for our healthcare than any

other high-income nation. Yet in 2015, over 32 million people in the United States, about 13.2%, were without insurance. Even today, however, several years after the Act was signed into law and after it was mostly upheld by the Supreme Court, a 2015 Kaiser Foundation poll found that 43% of likely voters viewed it unfavorably. Why is this?

The debate over the ACA and healthcare reform could take an entire textbook, but what this chapter will do is introduce the basics of insurance and the problems insurance companies face. It is these problems, and how insurance companies respond to them that, in part, explain the ACA.

Note:

Introduction to Information, Risk, and Insurance

In this chapter, you will learn about:

- The Problem of Imperfect Information and Asymmetric Information
- Insurance and Imperfect Information

Every purchase is based on a belief about the satisfaction that the good or service will provide. In turn, these beliefs are based on the information that the buyer has available. For many products, the information available to the buyer or the seller is imperfect or unclear, which can either make buyers regret past purchases or avoid making future ones.

This chapter discusses how imperfect and asymmetric information affect markets. The first module of the chapter discusses how asymmetric information affects markets for goods, labor, and financial capital. When buyers have less information about the quality of the good (for example, a gemstone) than sellers do, sellers may be tempted to mislead buyers. If a buyer cannot have at least some confidence in the quality of what is being purchased, then he will be reluctant or unwilling to purchase the products. Thus, mechanisms are needed to bridge this information gap, so buyers and sellers can engage in a transaction.

The second module of the chapter discusses insurance markets, which also face similar problems of imperfect information. For example, a car insurance company would prefer to sell insurance only to those who are unlikely to have auto accidents—but it is hard for the firm to identify those perfectly safe drivers. Conversely, buyers of car insurance would like to persuade the auto insurance company that they are safe drivers and should pay only a low price for insurance. If insurance markets cannot find ways to grapple with these problems of imperfect information, then even people who have low or average risks of making claims may not be able to purchase insurance. The chapter on financial markets (markets for stocks and bonds) will show that the problems of imperfect information can be especially poignant. Imperfect information cannot be eliminated, but it can often be managed.

The Problem of Imperfect Information and Asymmetric Information

By the end of this section, you will be able to:

- Analyze the impact of both imperfect information and asymmetric information
- Evaluate the role of advertisements in creating imperfect information
- Identify ways to reduce the risk of imperfect information
- Explain how imperfect information can affect price, quantity, and quality

Consider a purchase that many people make at important times in their lives: buying expensive jewelry. In May 1994, Doree Lynn bought an expensive ring from a jeweler in Washington, D.C., which included an emerald that cost \$14,500. Several years later, the emerald fractured. Lynn took it to another jeweler who found that cracks in the emerald had been filled with an epoxy resin. Lynn sued the original jeweler in 1997 for selling her a treated emerald without telling her, and won. The case publicized a number of little-known facts about precious stones. Most emeralds have internal flaws, and so they are soaked in clear oil or an epoxy resin to hide the flaws and make the color more deep and clear. Clear oil can leak out over time, and epoxy resin can discolor with age or heat. However, using clear oil or epoxy to “fill” emeralds is completely legal, as long as it is disclosed.

After Doree Lynn’s lawsuit, the NBC news show “Dateline” bought emeralds at four prominent jewelry stores in New York City in 1997. All the sales clerks at these stores, unaware that they were being recorded on a hidden camera, said the stones were untreated. When the emeralds were tested at a laboratory, however, it was discovered they had all been treated with oil or epoxy. Emeralds are not the only gemstones that are treated. Diamonds, topaz, and tourmaline are also often irradiated to enhance colors. The general rule is that all treatments to gemstones should be revealed, but often disclosure is not made. As such, many buyers face a situation of **asymmetric information**, where the both parties involved in an economic transaction have an unequal amount of information (one party knows much more than the other).

Many economic transactions are made in a situation of **imperfect information**, where either the buyer, the seller, or both, are less than 100% certain about the qualities of what is being bought and sold. Also, the transaction may be characterized by asymmetric information, in which one party has more information than the other regarding the economic transaction. Let's begin with some examples of how imperfect information complicates transactions in goods, labor, and financial capital markets. The presence of imperfect information can easily cause a decline in prices or quantities of products sold. However, buyers and sellers also have incentives to create mechanisms that will allow them to make mutually beneficial transactions even in the face of imperfect information.

If you are unclear about the difference between asymmetric information and imperfect information, read the following Clear It Up feature.

Note:

What is the difference between imperfect and asymmetric information?

For a market to reach equilibrium sellers and buyers must have full information about the product's price and quality. If there is limited information, then buyers and sellers may not be able to transact or will possibly make poor decisions.

Imperfect information refers to the situation where buyers and/or sellers do not have all of the necessary information to make an informed decision about the price or quality of a product. The term imperfect information simply means that not all the information necessary to make an informed decision is known to the buyers and/or sellers. Asymmetric information is the condition where one party, either the buyer or the seller, has more information about the quality or price of the product than the other party. In either case (imperfect or asymmetric information) buyers or sellers need remedies to make more informed decisions.

“Lemons” and Other Examples of Imperfect Information

Consider Marvin, who is trying to decide whether to buy a used car. Let's assume that Marvin is truly clueless about what happens inside a car's engine. He is willing to do some background research, like reading *Consumer Reports* or checking websites that offer information about makes and models of used cars and what they should cost. He might pay a mechanic to inspect the car. Even after devoting some money and time collecting information, however, Marvin still cannot be absolutely sure that he is buying a high-quality used car. He knows that he might buy the car, drive it home, and use it for a few weeks before discovering that car is a "lemon," which is slang for a defective product (especially a car).

Imagine that Marvin shops for a used car and finds two that look very similar in terms of mileage, exterior appearances, and age. One car costs \$4,000, while the other car costs \$4,600. Which car should Marvin buy?

If Marvin was choosing in a world of perfect information, the answer would be simple: he should buy the cheaper car. But Marvin is operating in a world of imperfect information, where the sellers likely know more about the car's problems than he does, and have an incentive to hide the information. After all, the more problems that are disclosed, the lower the car's selling price.

What should Marvin do? First, he needs to understand that even with imperfect information, prices still reflect information. Typically, used cars are more expensive on some dealer lots because the dealers have a trustworthy reputation to uphold. Those dealers try to fix problems that may not be obvious to their customers, in order to create good word of mouth about their vehicles' long term reliability. The short term benefits of selling their customers a "lemon" could cause a quick collapse in the dealer's reputation and a loss of long term profits. On other lots that are less well-established, one can find cheaper used cars, but the buyer takes on more risk when a dealer's reputation has little at stake. The cheapest cars of all often appear on Craigslist, where the individual seller has no reputation to defend. In sum, cheaper prices do carry more risk, so Marvin should balance his appetite for risk versus the potential headaches of many more unanticipated trips to the repair shop.

Similar problems with imperfect information arise in labor and financial capital markets. Consider Greta, who is applying for a job. Her potential employer, like the used car buyer, is concerned about ending up with a “lemon”—in this case a poor quality employee. The employer will collect information about Greta’s academic and work history. In the end, however, a degree of uncertainty will inevitably remain regarding Greta’s abilities, which are hard to demonstrate without actually observing her on the job. How can a potential employer screen for certain attributes, such as motivation, timeliness, ability to get along with others, and so on? Employers often look to trade schools and colleges to pre-screen candidates. Employers may not even interview a candidate unless he has a degree and, sometimes, a degree from a particular school. Employers may also view awards, a high grade point average, and other accolades as a signal of hard work, perseverance, and ability. Employers may also seek references for insights into key attributes such as energy level, work ethic, and so on.

How Imperfect Information Can Affect Equilibrium Price and Quantity

The presence of imperfect information can discourage both buyers and sellers from participating in the market. Buyers may become reluctant to participate because they cannot determine the quality of a product. Sellers of high-quality or medium-quality goods may be reluctant to participate, because it is difficult to demonstrate the quality of their goods to buyers—and since buyers cannot determine which goods have higher quality, they are likely to be unwilling to pay a higher price for such goods.

A market with few buyers and few sellers is sometimes referred to as a thin market. By contrast, a market with many buyers and sellers is called a thick market. When imperfect information is severe and buyers and sellers are discouraged from participating, markets may become extremely thin as a relatively small number of buyer and sellers attempt to communicate enough information that they can agree on a price.

When Price Mixes with Imperfect Information about Quality

A buyer confronted with imperfect information will often believe that the price being charged reveals something about the quality of the product. For example, a buyer may assume that a gemstone or a used car that costs more must be of higher quality, even though the buyer is not an expert on gemstones. Think of the expensive restaurant where the food must be good because it is so expensive or the shop where the clothes must be stylish because they cost so much, or the gallery where the art must be great, because it costs so much. If you are hiring a lawyer, you might assume that a lawyer who charges \$400 per hour must be better than a lawyer who charges \$150 per hour. In these cases, price can act as a signal of quality.

When buyers use the market price to draw inferences about the quality of products, then markets may have trouble reaching an equilibrium price and quantity. Imagine a situation where a used car dealer has a lot full of used cars that do not seem to be selling, and so the dealer decides to cut the prices of the cars to sell a greater quantity. In a market with imperfect information, many buyers may assume that the lower price implies low-quality cars. As a result, the lower price may not attract more customers. Conversely, a dealer who raises prices may find that customers assume that the higher price means that cars are of higher quality; as a result of raising prices, the dealer might sell more cars. (Whether or not consumers always behave rationally, as an economist would see it, is the subject of the following Clear It Up feature.)

The idea that higher prices might cause a greater quantity demanded and that lower prices might cause a lower quantity demanded runs exactly counter to the basic model of demand and supply (as outlined in the [Demand and Supply](#) chapter). These contrary effects, however, will reach natural limits. At some point, if the price is high enough, the quantity demanded will decline. Conversely, when the price declines far enough, buyers will increasingly find value even if the quality is lower. In addition, information eventually becomes more widely known. An overpriced restaurant that charges more than the quality of its food is worth to many buyers will not last forever.

Note:

Is consumer behavior rational?

There is a lot of human behavior out there that mainstream economists have tended to call “irrational” since it is consistently at odds with economists’ utility maximizing models. The typical response is for economists to brush these behaviors aside and call them “anomalies” or unexplained quirks.

“If only you knew more economics, you would not be so irrational,” is what many mainstream economists seem to be saying. A group known as behavioral economists has challenged this notion, because so much of this so-called “quirky” behavior is extremely common among us. For example, a conventional economist would say that if you lost a \$10 bill today, and also got an extra \$10 in your paycheck, you should feel perfectly neutral. After all, $-\$10 + \$10 = \$0$. You are the same financially as you were before. However, behavioral economists have done research that shows many people will feel some negative emotion—anger, frustration, and so forth—after those two things happen. We tend to focus more on the loss than the gain. This is known as “loss aversion,” where a \$1 loss pains us 2.25 times more than a \$1 gain helps us, according to the economists Daniel Kahneman and Amos Tversky in a famous 1979 *Econometrica* paper. This has implications for investing, as people tend to “overplay” the stock market by reacting more to losses than to gains.

Behavioral economics also tries to explain why people make seemingly irrational decisions in the presence of different situations, or how the decision is “framed.” A popular example is outlined here: Imagine you have the opportunity to buy an alarm clock for \$20 in Store A. Across the street, you learn, is the exact same clock at Store B for \$10. You might say it is worth your time—a five minute walk—to save \$10. Now, take a different example: You are in Store A buying a \$300 phone. Five minutes away, at Store B, the same phone is \$290. You again save \$10 by taking a five minute walk. Do you do it?

Surprisingly, it is likely that you would not. Mainstream economists would say “\$10 is \$10” and that it would be irrational to make a five minute walk for \$10 in one case and not the other. However, behavioral economists have pointed out that most of us evaluate outcomes relative to a reference point—here the cost of the product—and think of gains and losses as percentages rather than using actual savings.

Which view is right? Both have their advantages, but behavioral economists have at least shed a light on trying to describe and explain systematic behavior which previously has been dismissed as irrational. If most of us are engaged in some “irrational behavior,” perhaps there are deeper underlying reasons for this behavior in the first place.

Mechanisms to Reduce the Risk of Imperfect Information

If you were selling a good like emeralds or used cars where imperfect information is likely to be a problem, how could you reassure possible buyers? If you were buying a good where imperfect information is a problem, what would it take to reassure you? Buyers and sellers in the goods market rely on reputation as well as guarantees, warranties, and service contracts to assure product quality; in the labor market, occupational licenses and certifications are used to assure competency, while in financial capital market cosigners and collateral are used as insurance against unforeseen, detrimental events.

In the goods market, the seller of a good might offer a **money-back guarantee**, an agreement that functions as a promise of quality. This strategy may be especially important for a company that sells goods through mail-order catalogs or over the web, whose customers cannot see the actual products, because it encourages people to buy something even if they are not certain they want to keep it.

L.L. Bean started using money-back-guarantees in 1911, when the founder stitched waterproof shoe rubbers together with leather shoe tops, and sold them as hunting shoes. He guaranteed satisfaction. However, the stitching came apart and, out of the first batch of 100 pairs that were sold, 90 pairs were returned. L.L. Bean took out a bank loan, repaired all of the shoes, and replaced them. The L.L. Bean reputation for customer satisfaction began to spread. Many firms today offer money-back-guarantees for a few weeks or months, but L.L. Bean offers a complete money-back guarantee. Anything you have bought from L.L. Bean can always be returned, no matter how many years later or what condition the product is in, for a full money-back guarantee.

L.L. Bean has very few stores. Instead, most of its sales are made by mail, telephone, or, now, through their website. For this kind of firm, imperfect information may be an especially difficult problem, because customers cannot see and touch what they are buying. A combination of a money-back guarantee and a reputation for quality can help for a mail-order firm to flourish.

Sellers may offer a **warranty**, which is a promise to fix or replace the good, at least for a certain period of time. The seller may also offer a buyer a chance to buy a **service contract**, where the buyer pays an extra amount and the seller agrees to fix anything that goes wrong for a set time period. Service contracts are often used with large purchases such as cars, appliances and even houses.

Guarantees, warranties, and service contracts are examples of explicit reassurance that sellers provide. In many cases, firms also offer unstated guarantees. For example, some movie theaters might refund the cost of a ticket to a customer who walks out complaining about the show. Likewise, while restaurants do not generally advertise a money-back guarantee or exchange policies, many restaurants allow customers to exchange one dish for another or reduce the price of the bill if the customer is not satisfied.

The rationale for these policies is that firms want repeat customers, who in turn will recommend the business to others; as such, establishing a good reputation is of paramount importance. When buyers know that a firm is concerned about its reputation, they are less likely to worry about receiving a poor-quality product. For example, a well-established grocery store with a good reputation can often charge a higher price than a temporary stand at a local farmer's market, where the buyer may never see the seller again.

Sellers of labor provide information through resumes, recommendations, school transcripts, and examples of their work. **Occupational licenses** are also used to establish quality in the labor market. Occupational licenses, which are typically issued by government agencies, show that a worker has completed a certain type of education or passed a certain test. Some of the professionals who must hold a license are doctors, teachers, nurses, engineers, accountants, and lawyers. In addition, most states require a license to work as a barber, an embalmer, a dietitian, a massage therapist, a

hearing aid dealer, a counselor, an insurance agent, and a real estate broker. Some other jobs require a license in only one state. Minnesota requires a state license to be a field archeologist. North Dakota has a state license for bait retailers. In Louisiana, a state license is needed to be a “stress analyst” and California requires a state license to be a furniture upholsterer. According to a 2013 study from the University of Chicago, about 29% of U.S. workers have jobs that require occupational licenses.

Occupational licenses have their downside as well, as they represent a barrier to entry to certain industries. This makes it more difficult for new entrants to compete with incumbents, which can lead to higher prices and less consumer choice. In industries that require licenses, the government has decided that the additional information provided by licenses outweighs the negative effect on competition.

Note:

Are advertisers allowed to benefit from imperfect information?

Many advertisements seem full of imperfect information—at least by what they imply. Driving a certain car, drinking a particular soda, or wearing a certain shoe are all unlikely to bring fashionable friends and fun automatically, if at all. The government rules on advertising, enforced by the Federal Trade Commission (FTC), allow advertising to contain a certain amount of exaggeration about the general delight of using a product. They, however, also demand that if a claim is presented as a fact, it must be true.

Legally, deceptive advertising dates back to the 1950s when Colgate-Palmolive created a television advertisement that seemed to show Rapid Shave shaving cream being spread on sandpaper and then the sand was shaved off the sandpaper. What the television advertisement actually showed was sand sprinkled on Plexiglas—without glue—and then scraped aside by the razor.

In the 1960s, in magazine advertisements for Campbell’s vegetable soup, the company was having problems getting an appetizing picture of the soup, because the vegetables kept sinking. So they filled a bowl with

marbles and poured the soup over the top, so that the bowl appeared to be crammed with vegetables.

In the late 1980s, the Volvo Company filmed a television advertisement that showed a monster truck driving over cars, crunching their roofs—all except for the Volvo, which did not crush. However, the FTC found in 1991 that the roof of the Volvo used in the filming had been reinforced with an extra steel framework, while the roof supports on the other car brands had been cut.

The Wonder Bread Company ran television advertisements featuring “Professor Wonder,” who said that because Wonder Bread contained extra calcium, it would help children’s minds work better and improve their memory. The FTC objected, and in 2002 the company agreed to stop running the advertisements.

As can be seen in each of these cases, factual claims about the product’s performance are often checked, at least to some extent, by the Federal Trade Commission. Language and images that are exaggerated or ambiguous, but not actually false, are allowed in advertising. Untrue “facts” are not allowed. In any case, an old Latin saying applies when watching advertisements: *Caveat emptor*—that is, “let the buyer beware.”

On the buyer’s side of the labor market, a standard precaution against hiring a “lemon” of an employee is to specify that the first few months of employment are officially a trial or probationary period, and that the worker can be let go for any reason or no reason after that time. Sometimes workers also receive lower pay during this trial period.

In the financial capital market, before a bank makes a loan, it requires a prospective borrower fill out forms regarding the sources of income; in addition, the bank conducts a credit check on the individual’s past borrowing. Another approach is to require a **cosigner** on a loan; that is, another person or firm who legally pledges to repay some or all of the money if the original borrower does not do so. Yet another approach is to require **collateral**, often property or equipment that the bank would have a right to seize and sell if the loan is not repaid.

Buyers of goods and services cannot possibly become experts in evaluating the quality of gemstones, used cars, lawyers, and everything else they buy. Employers and lenders cannot be perfectly omniscient about whether possible workers will turn out well or potential borrowers will repay loans on time. But the mechanisms mentioned above can reduce the risks associated with imperfect information so that the buyer and seller are willing to proceed.

Key Concepts and Summary

Many economic transactions are made in a situation of imperfect information, where either the buyer, the seller, or both are less than 100% certain about the qualities of what is being bought and sold. When information about the quality of products is highly imperfect, it may be difficult for a market to exist.

A “lemon” is the name given to a product that turns out, after the purchase, to have low quality. When the seller has more accurate information about the quality of the product than the buyer, the buyer will be hesitant to buy, out of fear of purchasing a “lemon.”

Markets have many ways to deal with imperfect information. In goods markets, buyers facing imperfect information about products may depend upon money-back guarantees, warranties, service contracts, and reputation. In labor markets, employers facing imperfect information about potential employees may turn to resumes, recommendations, occupational licenses for certain jobs, and employment for trial periods. In capital markets, lenders facing imperfect information about borrowers may require detailed loan applications and credit checks, cosigners, and collateral.

Self-Check Questions

Exercise:

Problem:

For each of the following purchases, say whether you would expect the degree of imperfect information to be relatively high or relatively low:

- a. Buying apples at a roadside stand
- b. Buying dinner at the neighborhood restaurant around the corner
- c. Buying a used laptop computer at a garage sale
- d. Ordering flowers over the Internet for your friend in a different city

Solution:

- a. Imperfect information is relatively low; after all, you can see the apples.
- b. Imperfect information is relatively low. The neighborhood restaurant probably has a certain local reputation.
- c. Imperfect information is relatively high. How can you tell whether the computer is really in good working order? Why are they selling it?
- d. Imperfect information is relatively high. What do those flowers really look like?

Exercise:**Problem:**

Why is there asymmetric information in the labor market? What signals can an employer look for that might indicate the traits they are seeking in a new employee?

Solution:

Asymmetric information often exists in the labor market because employers cannot observe many key employee attributes until after the person is hired. Employees, however, know whether they are energetic or detailed-oriented. Employers, therefore, often seek schools to pre-

screen candidates. Employers may not even interview a candidate unless he has a degree and often a degree from a particular school. Employers may also view awards, a high grade point average, and other accolades as a signal of hard work, perseverance, and ability. Finally, employers seek references for insights into key attributes such as energy level, work ethic, and so on.

Review Questions

Exercise:

Problem:

Why might it be difficult for a buyer and seller to agree on a price when imperfect information exists?

Exercise:

Problem:

What do economists (and used-car dealers) mean by a “lemon”?

Exercise:

Problem:

What are some of the ways a seller of goods might reassure a possible buyer who is faced with imperfect information?

Exercise:

Problem:

What are some of the ways a seller of labor (that is, someone looking for a job) might reassure a possible employer who is faced with imperfect information?

Exercise:

Problem:

What are some of the ways that someone looking for a loan might reassure a bank that is faced with imperfect information about whether the loan will be repaid?

Critical Thinking Questions**Exercise:****Problem:**

You are on the board of directors of a private high school, which is hiring new tenth-grade science teachers. As you think about hiring someone for a job, what are some mechanisms you might use to overcome the problem of imperfect information?

Exercise:**Problem:**

A website offers a place for people to buy and sell emeralds, but information about emeralds can be quite imperfect. The website then enacts a rule that all sellers in the market must pay for two independent examinations of their emerald, which are available to the customer for inspection.

- a. How would you expect this improved information to affect demand for emeralds on this website?
- b. How would you expect this improved information to affect the quantity of high-quality emeralds sold on the website?

Problem**Exercise:**

Problem:

Using [\[link\]](#), sketch the effects in parts (a) and (b) on a single supply and demand diagram. What prediction would you make about how the improved information alters the equilibrium quantity and price?

References

Center on Budget and Policy Priorities. 2015. "Policy Basics: Where Do Our Federal Tax Dollars Go?" Accessed April 1, 2015.
<http://www.cbpp.org/cms/?fa=view&id=1258>.

Consumer Reports. "Consumer Reports.org."
<http://www.consumerreports.org/cro/index.htm>.

Federal Trade Commission. "About the Federal Trade Commission." Last modified October 17, 2013. <http://www.ftc.gov/ftc/about.shtm>.

Huffington Post. 2015. "HUFFPOLLSTER: Poll Shows Uptick In Obamacare Favorable Rating." Accessed April 1, 2015.
http://www.huffingtonpost.com/2015/03/19/affordable-care-act-fav_n_6900938.html.

Kahneman, Daniel, and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*. 47, no. 2 (1979): 263-291.
http://www.princeton.edu/~kahneman/docs/Publications/prospect_theory.pdf.

Rasmussen Reports, LLC. "Rasmussen Reports."
<http://www.rasmussenreports.com/>.

Glossary

asymmetric information

a situation where the seller or the buyer has more information than the other regarding the quality of the item being sold

collateral

something valuable—often property or equipment—that a lender would have a right to seize and sell if the loan is not repaid

cosigner

another person or firm who legally pledges to repay some or all of the money on a loan if the original borrower does not do so

imperfect information

a situation where either the buyer or the seller, or both, are uncertain about the qualities of what is being bought and sold

money-back guarantee

a promise that the buyer's money will be refunded under certain conditions

occupational license

licenses issued by government agencies, which indicate that a worker has completed a certain type of education or passed a certain test

service contract

the buyer pays an extra amount and the seller agrees to fix anything specified in the contract that goes wrong for a set time period

warranty

a promise to fix or replace the good for a certain period of time

Insurance and Imperfect Information

By the end of this section, you will be able to:

- Explain how insurance works
- Identify and evaluate various forms of government and social insurance
- Discuss the problems caused by moral hazard and adverse selection
- Analyze the impact of government regulation of insurance

Insurance is a method that households and firms use to prevent any single event from having a significant detrimental financial effect. Generally, households or firms with insurance make regular payments, called **premiums**. The insurance company prices these premiums based on the probability of certain events occurring among a pool of people. Members of the group who then suffer a specified bad experience receive payments from this pool of money.

Many people have several kinds of insurance: health insurance that pays when they receive medical care; car insurance that pays if they are the driver in an automobile accident; house or renter's insurance that pays if possessions are stolen or damaged by fire; and life insurance, which pays for the family if the principal dies. [\[link\]](#) lists a set of insurance markets.

Type of Insurance	Who Pays for It?	It Pays Out When . . .
Health insurance	Employers and individuals	Medical expenses are incurred
Life insurance	Employers and individuals	Policyholder dies

Type of Insurance	Who Pays for It?	It Pays Out When . . .
Automobile insurance	Individuals	Car is damaged, stolen, or causes damage to others
Property and homeowner's insurance	Homeowners and renters	Dwelling is damaged or burglarized
Liability insurance	Firms and individuals	An injury occurs for which you are partly responsible
Malpractice insurance	Doctors, lawyers, and other professionals	A poor quality of service is provided that causes harm to others

Some Insurance Markets

All insurance involves imperfect information in both an obvious way and in a deeper way. At an obvious level, future events cannot be predicted with certainty. For example, it cannot be known with certainty who will have a car accident, become ill, die, or have his home robbed in the next year. Imperfect information also applies to estimating the risk that something will happen to any individual. It is difficult for an insurance company to estimate the risk that, say, a particular 20-year-old male driver from New York City will have an accident, because even within that group, some drivers will drive more safely than others. Thus, adverse events occur out of a combination of people's characteristics and choices that make the risks higher or lower and then the good or bad luck of what actually happens.

The major additional costs to insurance companies, other than the payment of claims, are the costs of running a business: the administrative costs of hiring workers, administering accounts, and processing insurance claims. For most insurance companies, the insurance premiums coming in and the

claims payments going out are much larger than the amounts earned by investing money or the administrative costs.

Thus, while factors like investment income earned on reserves, administrative costs, and groups with different risks complicate the overall picture, a fundamental law of insurance must hold true: The average person's payments into insurance over time must cover 1) the average person's claims, 2) the costs of running the company, and 3) leave room for the firm's profits. This law can be boiled down to the idea that average premiums and average insurance payouts must be approximately equal.

The Moral Hazard Problem

Moral hazard refers to the case when people engage in riskier behavior with insurance than they would if they did not have insurance. For example, if you have health insurance that covers the cost of visiting the doctor, you may be less likely to take precautions against catching an illness that might require a doctor's visit. If you have car insurance, you will worry less about driving or parking your car in ways that make it more likely to get dented. In another example, a business without insurance might install absolute top-level security and fire sprinkler systems to guard against theft and fire. If it is insured, that same business might only install a minimum level of security and fire sprinkler systems.

Moral hazard cannot be eliminated, but insurance companies have some ways of reducing its effect. Investigations to prevent insurance fraud are one way of reducing the extreme cases of moral hazard. Insurance companies can also monitor certain kinds of behavior; to return to the example from above, they might offer a business a lower rate on property insurance if the business installs a top-level security and fire sprinkler system and has those systems inspected once a year.

Another method to reduce moral hazard is to require the injured party to pay a share of the costs. For example, insurance policies often have **deductibles**, which is an amount that the insurance policyholder must pay out of their own pocket before the insurance coverage starts paying. For example, auto insurance might pay for all losses greater than \$500. Health

insurance policies often have a **copayment**, in which the policyholder must pay a small amount; for example, a person might have to pay \$20 for each doctor visit, and the insurance company would cover the rest. Another method of cost-sharing is **coinsurance**, which means that the insurance company covers a certain percentage of the cost. For example, insurance might pay for 80% of the costs of repairing a home after a fire, but the homeowner would pay the other 20%.

All of these forms of cost-sharing discourage moral hazard, because people know that they will have to pay something out of their own pocket when they make an insurance claim. The effect can be powerful. One prominent study found that when people face moderate deductibles and copayments for their health insurance, they consume about one-third less in medical care than people who have complete insurance and do not pay anything out of pocket, presumably because deductibles and copayments reduce the level of moral hazard. However, those who consumed less health care did not seem to have any difference in health status.

A final way of reducing moral hazard, which is especially applicable to health care, is to focus on the incentives of providers of health care, rather than consumers. Traditionally, most health care in the United States has been provided on a **fee-for-service** basis, which means that medical care providers are paid for the services they provide and are paid more if they provide additional services. However, in the last decade or so, the structure of healthcare provision has shifted to an emphasis on health maintenance organizations (HMOs). A **health maintenance organization (HMO)** provides healthcare that receives a fixed amount per person enrolled in the plan—regardless of how many services are provided. In this case, a patient with insurance has an incentive to demand more care, but the healthcare provider, which is receiving only a fixed payment, has an incentive to reduce the moral hazard problem by limiting the quantity of care provided—as long as it will not lead to worse health problems and higher costs later. Today, many doctors are paid with some combination of managed care and fee-for-service; that is, a flat amount per patient, but with additional payments for the treatment of certain health conditions.

Imperfect information is the cause of the moral hazard problem. If an insurance company had perfect information on risk, it could simply raise its premiums every time an insured party engages in riskier behavior. However, an insurance company cannot monitor all the risks that people take all the time and so, even with various checks and cost-sharing, moral hazard will remain a problem.

The Adverse Selection Problem

Adverse selection refers to the problem in which the buyers of insurance have more information about whether they are high-risk or low-risk than the insurance company does. This creates an asymmetric information problem for the insurance company because buyers who are high-risk tend to want to buy more insurance, without letting the insurance company know about their higher risk. For example, someone purchasing health insurance or life insurance probably knows more about their family's health history than an insurer can reasonably find out even with a costly investigation; someone purchasing car insurance may know that they are a high-risk driver who has not yet had a major accident—but it is hard for the insurance company to collect information about how people actually drive.

To understand how adverse selection can strangle an insurance market, recall the situation of 100 drivers who are buying automobile insurance, where 60 drivers had very low damages of \$100 each, 30 drivers had medium-sized accidents that cost \$1,000 each, and 10 of the drivers had large accidents that cost \$15,000. That would equal \$186,000 in total payouts by the insurance company. Imagine that, while the insurance company knows the overall size of the losses, it cannot identify the high-risk, medium-risk, and low-risk drivers. However, the drivers themselves know their risk groups. Since there is asymmetric information between the insurance company and the drivers, the insurance company would likely set the price of insurance at \$1,860 per year, to cover the average loss (not including the cost of overhead and profit). The result is that those with low risks of only \$100 will likely decide not to buy insurance; after all, it makes no sense for them to pay \$1,860 per year when they are likely only to experience losses of \$100. Those with medium risks of a \$1,000 accident will not buy insurance either. So the insurance company ends up only

selling insurance for \$1,860 to high-risk drivers who will average \$15,000 in claims apiece. So the insurance company ends up losing a lot of money. If the insurance company tries to raise its premiums to cover the losses of those with high risks, then those with low or medium risks will be even more discouraged from buying insurance.

Rather than face such a situation of adverse selection, the insurance company may decide not to sell insurance in this market at all. If an insurance market is to exist, then one of two things must happen. First, the insurance company might find some way of separating insurance buyers into risk groups with some degree of accuracy and charging them accordingly, which in practice often means that the insurance company tries not to sell insurance to those who may pose high risks. Or second, those with low risks must be required to buy insurance, even if they have to pay more than the actuarially fair amount for their risk group. The notion that people can be required to purchase insurance raises the issue of government laws and regulations that influence the insurance industry.

Key Concepts and Summary

Insurance is a way of sharing risk. A group of people pay premiums for insurance against some unpleasant event, and those in the group who actually experience the unpleasant event then receive some compensation. The fundamental law of insurance is that what the average person pays in over time must be very similar to what the average person gets out. In an actuarially fair insurance policy, the premiums that a person pays to the insurance company are the same as the average amount of benefits for a person in that risk group. Moral hazard arises in insurance markets because those who are insured against a risk will have less reason to take steps to avoid the costs from that risk.

Many insurance policies have deductibles, copayments, or coinsurance. A deductible is the maximum amount that the policyholder must pay out-of-pocket before the insurance company pays the rest of the bill. A copayment is a flat fee that an insurance policy-holder must pay before receiving services. Coinsurance requires the policyholder to pay a certain percentage of costs. Deductibles, copayments, and coinsurance reduce moral hazard by

requiring the insured party to bear some of the costs before collecting insurance benefits.

In a fee-for-service health financing system, medical care providers are reimbursed according to the cost of services they provide. An alternative method of organizing health care is through health maintenance organizations (HMOs), where medical care providers are reimbursed according to the number of patients they handle, and it is up to the providers to allocate resources between patients who receive more or fewer health care services. Adverse selection arises in insurance markets when insurance buyers know more about the risks they face than does the insurance company. As a result, the insurance company runs the risk that low-risk parties will avoid its insurance because it is too costly for them, while high-risk parties will embrace it because it looks like a good deal to them.

Self-Check Question

Exercise:

Problem: Why is it difficult to measure health outcomes?

Solution:

It is almost impossible to distinguish whether a health outcome such as life expectancy was the result of personal preferences that might affect health and longevity, such as diet, exercise, certain risky behavior, and consumption of certain items like tobacco, or the result of expenditures on health care (for example, annual check-ups).

Review Questions

Exercise:

Problem: What is an insurance premium?

Exercise:

Problem:

In an insurance system, would you expect each person to receive in benefits pretty much what they pay in premiums? Or is it just that the average benefits paid will equal the average premiums paid?

Exercise:

Problem: What is an actuarially fair insurance policy?

Exercise:

Problem: What is the problem of moral hazard?

Exercise:

Problem:

How can moral hazard lead to insurance being more costly than was expected?

Exercise:

Problem: Define deductibles, copayments, and coinsurance.

Exercise:

Problem:

How can deductibles, copayments, and coinsurance reduce moral hazard?

Exercise:

Problem:

What is the key difference between a fee-for-service healthcare system and a system based on health maintenance organizations?

Exercise:

Problem:

How might adverse selection make it difficult for an insurance market to operate?

Exercise:**Problem:**

What are some of the metrics used to measure health outcomes?

Critical Thinking Questions**Exercise:****Problem:**

How do you think the problem of moral hazard might have affected the safety of sports such as football and boxing when safety regulations started requiring that players wear more padding?

Exercise:**Problem:**

To what sorts of customers would an insurance company offer a policy with a high copay? What about a high premium with a lower copay?

References

Central Intelligence Agency. "The World Factbook."
<https://www.cia.gov/library/publications/the-world-factbook/index.html>.

National Association of Insurance Commissioners. "National Association of Insurance Commissioners & The Center for Insurance Policy and Research." <http://www.naic.org/>.

OECD. "The Organisation for Economic Co-operation and Development (OECD)." <http://www.oecd.org/about/>.

USA Today. 2015. “Uninsured Rates Drop Dramatically under Obamacare.” Accessed April 1, 2015.
<http://www.usatoday.com/story/news/nation/2015/03/16/uninsured-rates-drop-sharply-under-obamacare/24852325/>.

Thaler, Richard H., and Sendhil Mullainathan. “The Concise Encyclopedia of Economics: Behavioral Economics.” *Library of Economics and Liberty*.
<http://www.econlib.org/library/Enc/BehavioralEconomics.html>.

Henry J. Kaiser Family Foundation, The. “Health Reform: Summary of the Affordable care Act.” Last modified April 25, 2013. <http://kff.org/health-reform/fact-sheet/summary-of-new-health-reform-law/>.

Glossary

adverse selection

when groups with inherently higher risks than the average person seek out insurance, thus straining the insurance system

coinsurance

when an insurance policyholder pays a percentage of a loss, and the insurance company pays the remaining cost

copayment

when an insurance policyholder must pay a small amount for each service, before insurance covers the rest

deductible

an amount that the insurance policyholders must pay out of their own pocket before the insurance coverage pays anything

fee-for-service

when medical care providers are paid according to the services they provide

health maintenance organization (HMO)

an organization that provides health care and is paid a fixed amount per person enrolled in the plan—regardless of how many services are

provided

insurance

method of protecting a person from financial loss, whereby policy holders make regular payments to an insurance entity; the insurance firm then remunerates a group member who suffers significant financial damage from an event covered by the policy

moral hazard

when people have insurance against a certain event, they are less likely to guard against that event occurring

premium

payment made to an insurance company

risk group

a group that shares roughly the same risks of an adverse event occurring

Perfect Competition and Why It Matters

By the end of this section, you will be able to:

- Explain the characteristics of a perfectly competitive market
- Discuss how perfectly competitive firms react in the short run and in the long run

Firms are said to be in **perfect competition** when the following conditions occur: (1) many firms produce identical products; (2) many buyers are available to buy the product, and many sellers are available to sell the product; (3) sellers and buyers have all relevant information to make rational decisions about the product being bought and sold; and (4) firms can enter and leave the market without any restrictions—in other words, there is free entry and exit into and out of the market.

A perfectly competitive firm is known as a **price taker**, because the pressure of competing firms forces them to accept the prevailing equilibrium price in the market. If a firm in a perfectly competitive market raises the price of its product by so much as a penny, it will lose all of its sales to competitors. When a wheat grower, as discussed in the Bring it Home feature, wants to know what the going price of wheat is, he or she has to go to the computer or listen to the radio to check. The market price is determined solely by supply and demand in the entire market and not the individual farmer. Also, a perfectly competitive firm must be a very small player in the overall market, so that it can increase or decrease output without noticeably affecting the overall quantity supplied and price in the market.

A perfectly competitive market is a hypothetical extreme; however, producers in a number of industries do face many competitor firms selling highly similar goods, in which case they must often act as price takers. Agricultural markets are often used as an example. The same crops grown by different farmers are largely interchangeable. According to the United States Department of Agriculture monthly reports, in 2015, U.S. corn farmers received an average price of \$6.00 per bushel and wheat farmers received an average price of \$6.00 per bushel. A corn farmer who attempted to sell at \$7.00 per bushel, or a wheat grower who attempted to sell for \$8.00 per bushel, would not have found any buyers. A perfectly competitive

firm will not sell below the equilibrium price either. Why should they when they can sell all they want at the higher price? Other examples of agricultural markets that operate in close to perfectly competitive markets are small roadside produce markets and small organic farmers.

Note:

Visit this [website](#) that reveals the current value of various commodities.



This chapter examines how profit-seeking firms decide how much to produce in perfectly competitive markets. Such firms will analyze their costs as discussed in the chapter on [Cost and Industry Structure](#). In the short run, the perfectly competitive firm will seek the quantity of output where profits are highest or, if profits are not possible, where losses are lowest. In this example, the “short run” refers to a situation in which firms are producing with one fixed input and incur fixed costs of production. (In the real world, firms can have many fixed inputs.)

In the long run, perfectly competitive firms will react to profits by increasing production. They will respond to losses by reducing production or exiting the market. Ultimately, a long-run *equilibrium* will be attained when no new firms want to enter the market and existing firms do not want to leave the market, as economic profits have been driven down to zero.

Key Concepts and Summary

A perfectly competitive firm is a price taker, which means that it must accept the equilibrium price at which it sells goods. If a perfectly competitive firm attempts to charge even a tiny amount more than the market price, it will be unable to make any sales. In a perfectly competitive market there are thousands of sellers, easy entry, and identical products. A short-run production period is when firms are producing with some fixed inputs. Long-run equilibrium in a perfectly competitive industry occurs after all firms have entered and exited the industry and seller profits are driven to zero.

Perfect competition means that there are many sellers, there is easy entry and exiting of firms, products are identical from one seller to another, and sellers are price takers.

Self-Check Questions

Exercise:

Problem:

Firms in a perfectly competitive market are said to be “price takers”—that is, once the market determines an equilibrium price for the product, firms must accept this price. If you sell a product in a perfectly competitive market, but you are not happy with its price, would you raise the price, even by a cent?

Solution:

No, you would not raise the price. Your product is exactly the same as the product of the many other firms in the market. If your price is greater than that of your competitors, then your customers would switch to them and stop buying from you. You would lose all your sales.

Exercise:

Problem:

Would independent trucking fit the characteristics of a perfectly competitive industry?

Solution:

Possibly. Independent truckers are by definition small and numerous. All that is required to get into the business is a truck (not an inexpensive asset, though) and a commercial driver's license. To exit, one need only sell the truck. All trucks are essentially the same, providing transportation from point A to point B. (We're assuming we not talking about specialized trucks.) Independent truckers must take the going rate for their service, so independent trucking does seem to have most of the characteristics of perfect competition.

Review Questions**Exercise:****Problem:**

A single firm in a perfectly competitive market is relatively small compared to the rest of the market. What does this mean? How "small" is "small"?

Exercise:**Problem:**

What are the four basic assumptions of perfect competition? Explain in words what they imply for a perfectly competitive firm.

Exercise:

Problem: What is a "price taker" firm?

Critical Thinking Questions

Exercise:

Problem:

Finding a life partner is a complicated process that may take many years. It is hard to think of this process as being part of a very complex market, with a demand and a supply for partners. Think about how this market works and some of its characteristics, such as search costs. Would you consider it a perfectly competitive market?

Exercise:

Problem:

Can you name five examples of perfectly competitive markets? Why or why not?

Glossary

market structure

the conditions in an industry, such as number of sellers, how easy or difficult it is for a new firm to enter, and the type of products that are sold

perfect competition

each firm faces many competitors that sell identical products

price taker

a firm in a perfectly competitive market that must take the prevailing market price as given

Perfect Competition and Why It Matters

By the end of this section, you will be able to:

- Explain the characteristics of a perfectly competitive market
- Discuss how perfectly competitive firms react in the short run and in the long run

Firms are said to be in **perfect competition** when the following conditions occur: (1) many firms produce identical products; (2) many buyers are available to buy the product, and many sellers are available to sell the product; (3) sellers and buyers have all relevant information to make rational decisions about the product being bought and sold; and (4) firms can enter and leave the market without any restrictions—in other words, there is free entry and exit into and out of the market.

A perfectly competitive firm is known as a **price taker**, because the pressure of competing firms forces them to accept the prevailing equilibrium price in the market. If a firm in a perfectly competitive market raises the price of its product by so much as a penny, it will lose all of its sales to competitors. When a wheat grower, as discussed in the Bring it Home feature, wants to know what the going price of wheat is, he or she has to go to the computer or listen to the radio to check. The market price is determined solely by supply and demand in the entire market and not the individual farmer. Also, a perfectly competitive firm must be a very small player in the overall market, so that it can increase or decrease output without noticeably affecting the overall quantity supplied and price in the market.

A perfectly competitive market is a hypothetical extreme; however, producers in a number of industries do face many competitor firms selling highly similar goods, in which case they must often act as price takers. Agricultural markets are often used as an example. The same crops grown by different farmers are largely interchangeable. According to the United States Department of Agriculture monthly reports, in 2015, U.S. corn farmers received an average price of \$6.00 per bushel and wheat farmers received an average price of \$6.00 per bushel. A corn farmer who attempted to sell at \$7.00 per bushel, or a wheat grower who attempted to sell for \$8.00 per bushel, would not have found any buyers. A perfectly competitive

firm will not sell below the equilibrium price either. Why should they when they can sell all they want at the higher price? Other examples of agricultural markets that operate in close to perfectly competitive markets are small roadside produce markets and small organic farmers.

Key Concepts and Summary

A perfectly competitive firm is a price taker, which means that it must accept the equilibrium price at which it sells goods. If a perfectly competitive firm attempts to charge even a tiny amount more than the market price, it will be unable to make any sales. In a perfectly competitive market there are thousands of sellers, easy entry, and identical products. A short-run production period is when firms are producing with some fixed inputs. Long-run equilibrium in a perfectly competitive industry occurs after all firms have entered and exited the industry and seller profits are driven to zero.

Perfect competition means that there are many sellers, there is easy entry and exiting of firms, products are identical from one seller to another, and sellers are price takers.

Self-Check Questions

Exercise:

Problem:

Firms in a perfectly competitive market are said to be “price takers”—that is, once the market determines an equilibrium price for the product, firms must accept this price. If you sell a product in a perfectly competitive market, but you are not happy with its price, would you raise the price, even by a cent?

Solution:

No, you would not raise the price. Your product is exactly the same as the product of the many other firms in the market. If your price is

greater than that of your competitors, then your customers would switch to them and stop buying from you. You would lose all your sales.

Exercise:

Problem:

Would independent trucking fit the characteristics of a perfectly competitive industry?

Solution:

Possibly. Independent truckers are by definition small and numerous. All that is required to get into the business is a truck (not an inexpensive asset, though) and a commercial driver's license. To exit, one need only sell the truck. All trucks are essentially the same, providing transportation from point A to point B. (We're assuming we not talking about specialized trucks.) Independent truckers must take the going rate for their service, so independent trucking does seem to have most of the characteristics of perfect competition.

Review Questions

Exercise:

Problem:

A single firm in a perfectly competitive market is relatively small compared to the rest of the market. What does this mean? How "small" is "small"?

Exercise:

Problem:

What are the four basic assumptions of perfect competition? Explain in words what they imply for a perfectly competitive firm.

Exercise:

Problem: What is a “price taker” firm?

Critical Thinking Questions

Exercise:

Problem:

Finding a life partner is a complicated process that may take many years. It is hard to think of this process as being part of a very complex market, with a demand and a supply for partners. Think about how this market works and some of its characteristics, such as search costs. Would you consider it a perfectly competitive market?

Exercise:

Problem:

Can you name five examples of perfectly competitive markets? Why or why not?

Glossary

market structure

the conditions in an industry, such as number of sellers, how easy or difficult it is for a new firm to enter, and the type of products that are sold

perfect competition

each firm faces many competitors that sell identical products

price taker

a firm in a perfectly competitive market that must take the prevailing market price as given

Efficiency in Perfectly Competitive Markets

By the end of this section, you will be able to:

- Apply concepts of productive efficiency and allocative efficiency to perfectly competitive markets
- Compare the model of perfect competition to real-world markets

When profit-maximizing firms in perfectly competitive markets combine with utility-maximizing consumers, something remarkable happens: the resulting quantities of outputs of goods and services demonstrate both productive and allocative efficiency (terms that were first introduced in [Choice in a World of Scarcity](#)) .

Productive efficiency means producing without waste, so that the choice is on the production possibility frontier. In the long run in a perfectly competitive market, because of the process of entry and exit, the price in the market is equal to the minimum of the long-run average cost curve. In other words, goods are being produced and sold at the lowest possible average cost.

Allocative efficiency means that among the points on the production possibility frontier, the point that is chosen is socially preferred—at least in a particular and specific sense. In a perfectly competitive market, price will be equal to the marginal cost of production. Think about the price that is paid for a good as a measure of the social benefit received for that good; after all, willingness to pay conveys what the good is worth to a buyer. Then think about the marginal cost of producing the good as representing not just the cost for the firm, but more broadly as the social cost of producing that good. When perfectly competitive firms follow the rule that profits are maximized by producing at the quantity where price is equal to marginal cost, they are thus ensuring that the social benefits received from producing a good are in line with the social costs of production.

The statements that a perfectly competitive market in the long run will feature both productive and allocative efficiency do need to be taken with a few grains of salt. Remember, economists are using the concept of “efficiency” in a particular and specific sense, not as a synonym for “desirable in every way.” For one thing, consumers’ ability to pay reflects

the income distribution in a particular society. Thus, a homeless person may have no ability to pay for housing because they have insufficient income.

Perfect competition, in the long run, is a hypothetical benchmark. For market structures such as monopoly, monopolistic competition, and oligopoly, which are more frequently observed in the real world than perfect competition, firms will not always produce at the minimum of average cost, nor will they always set price equal to marginal cost. Thus, these other competitive situations will not produce productive and allocative efficiency.

Moreover, real-world markets include many issues that are assumed away in the model of perfect competition, including pollution, inventions of new technology, poverty which may make some people unable to pay for basic necessities of life, government programs like national defense or education, discrimination in labor markets, and buyers and sellers who must deal with imperfect and unclear information. These issues are explored in other chapters. However, the theoretical efficiency of perfect competition does provide a useful benchmark for comparing the issues that arise from these real-world problems.

Key Concepts and Summary

Long-run equilibrium in perfectly competitive markets meets two important conditions: allocative efficiency and productive efficiency. These two conditions have important implications. First, resources are allocated to their best alternative use. Second, they provide the maximum satisfaction attainable by society.

Self-Check Questions

Exercise:

Problem:

Productive efficiency and allocative efficiency are two concepts achieved in the long run in a perfectly competitive market. These are the two reasons why we call them “perfect.” How would you use these two concepts to analyze other market structures and label them “imperfect?”

Solution:

Perfect competition is considered to be “perfect” because both allocative and productive efficiency are met at the same time in a long-run equilibrium. If a market structure results in long-run equilibrium that does not minimize average total costs and/or does not charge a price equal to marginal cost, then either allocative or productive (or both) efficiencies are not met, and therefore the market cannot be labeled “perfect.”

Exercise:**Problem:**

Explain how the profit-maximizing rule of setting $P = MC$ leads a perfectly competitive market to be allocatively efficient.

Solution:

Think of the market price as representing the gain to society from a purchase, since it represents what someone is willing to pay. Think of the marginal cost as representing the cost to society from making the last unit of a good. If $P > MC$, then the benefits from producing more of a good exceed the costs, and society would gain from producing more of the good. If $P < MC$, then the social costs of producing the marginal good exceed the social benefits, and society should produce less of the good. Only if $P = MC$, the rule applied by a profit-maximizing perfectly competitive firm, will society’s costs and benefits be in balance. This choice will be the option that brings the greatest overall benefit to society.

Review Questions

Exercise:

Problem:

Will a perfectly competitive market display productive efficiency?
Why or why not?

Exercise:

Problem:

Will a perfectly competitive market display allocative efficiency? Why or why not?

Critical Thinking Questions

Exercise:

Problem:

Assuming that the market for cigarettes is in perfect competition, what does allocative and productive efficiency imply in this case? What does it not imply?

Exercise:

Problem:

In the argument for why perfect competition is allocatively efficient, the price that people are willing to pay represents the gains to society and the marginal cost to the firm represents the costs to society. Can you think of some social costs or issues that are not included in the marginal cost to the firm? Or some social gains that are not included in what people pay for a good?

References

Index Mundi. n.d. "Wheat Monthly Price—U.S. Dollars per Metric Ton." Accessed March 11, 2015. <http://www.indexmundi.com/commodities/?commodity=wheat>.

Knutson, J. "Wheat on the Defensive in the Northern Plains." *Agweek, Associated Press State Wire: North Dakota (ND)*. April 14, 2013.

SBA Office of Advocacy. 2014. "Frequently Asked Questions: Advocacy: the voice of small business in government." Accessed March 11, 2015. https://www.sba.gov/sites/default/files/advocacy/FAQ_March_2014_0.pdf.

Introduction to a Monopoly

class="introduction"

Political Power from a Cotton Monopoly

In the mid-nineteenth century, the United States, specifically the Southern states, had a near monopoly in the cotton supplied to Great Britain. These states attempted to leverage this economic power into political power—trying to sway Great Britain to formally recognize the Confederate States of America. (Credit: modification of work by “ashleylovespizza”/Flickr or Creative Commons)

**Note:****The Rest is History**

Many of the opening case studies have focused on current events. This one steps into the past to observe how monopoly, or near monopolies, have helped shape history. In the spring of 1773, the East India Company, a firm that, in its time, was designated ‘too big to fail,’ was continuing to experience financial difficulties. To help shore up the failing firm, the British Parliament authorized the Tea Act. The act continued the tax on teas and made the East India Company the sole legal supplier of tea to the American colonies. By November, the citizens of Boston had had enough. They refused to permit the tea to be unloaded, citing their main complaint: “No taxation without representation.” Arriving tea-bearing ships were warned via several newspapers, including *The Massachusetts Gazette*, “We are prepared, and shall not fail to pay them an unwelcome visit; by The Mohawks.”

Step forward in time to 1860—the eve of the American Civil War—to another near monopoly supplier of historical significance: the U.S. cotton industry. At that time, the Southern states provided the majority of the

cotton Britain imported. The South, wanting to secede from the Union, hoped to leverage Britain's high dependency on its cotton into formal diplomatic recognition of the Confederate States of America.

This leads us to the topic of this chapter: a firm that controls all (or nearly all) of the supply of a good or service—a monopoly. How do monopoly firms behave in the marketplace? Do they have “power?” Does this power potentially have unintended consequences? We'll return to this case at the end of the chapter to see how the tea and cotton monopolies influenced U.S. history.

Note:

Introduction to a Monopoly

In this chapter, you will learn about:

- How Monopolies form: Barriers to Entry
- How a Profit-Maximizing Monopoly Chooses Output and Price

There is a widespread belief that top executives at firms are the strongest supporters of market competition, but this belief is far from the truth. Think about it this way: If you very much wanted to win an Olympic gold medal, would you rather be far better than everyone else, or locked in competition with many athletes just as good as you are? Similarly, if you would like to attain a very high level of profits, would you rather manage a business with little or no competition, or struggle against many tough competitors who are trying to sell to your customers? By now, you might have read the chapter on [Perfect Competition](#). In this chapter, we explore the opposite extreme: monopoly.

If perfect competition is a market where firms have no market power and they simply respond to the market price, monopoly is a market with no competition at all, and firms have complete market power. In the case of **monopoly**, one firm produces all of the output in a market. Since a monopoly faces no significant competition, it can charge any price it

wishes. While a monopoly, by definition, refers to a single firm, in practice the term is often used to describe a market in which one firm merely has a very high market share. This tends to be the definition that the U.S. Department of Justice uses.

Even though there are very few true monopolies in existence, we do deal with some of those few every day, often without realizing it: The U.S. Postal Service, your electric and garbage collection companies are a few examples. Some new drugs are produced by only one pharmaceutical firm—and no close substitutes for that drug may exist.

From the mid-1990s until 2004, the U.S. Department of Justice prosecuted the Microsoft Corporation for including Internet Explorer as the default web browser with its operating system. The Justice Department's argument was that, since Microsoft possessed an extremely high market share in the industry for operating systems, the inclusion of a free web browser constituted unfair competition to other browsers, such as Netscape Navigator. Since nearly everyone was using Windows, including Internet Explorer eliminated the incentive for consumers to explore other browsers and made it impossible for competitors to gain a foothold in the market. In 2013, the Windows system ran on more than 90% of the most commonly sold personal computers. In 2015, a U.S. federal court tossed out antitrust charges that Google had an agreement with mobile device makers to set Google as the default search engine.

This chapter begins by describing how monopolies are protected from competition, including laws that prohibit competition, technological advantages, and certain configurations of demand and supply. It then discusses how a monopoly will choose its profit-maximizing quantity to produce and what price to charge. While a monopoly must be concerned about whether consumers will purchase its products or spend their money on something altogether different, the monopolist need not worry about the actions of other competing firms producing its products. As a result, a monopoly is not a price taker like a perfectly competitive firm, but instead exercises some power to choose its market price.

How Monopolies Form: Barriers to Entry

By the end of this section, you will be able to:

- Distinguish between a natural monopoly and a legal monopoly.
- Explain how economies of scale and the control of natural resources led to the necessary formation of legal monopolies
- Analyze the importance of trademarks and patents in promoting innovation
- Identify examples of predatory pricing

Because of the lack of competition, monopolies tend to earn significant economic profits. These profits should attract vigorous competition as described in [Perfect Competition](#), and yet, because of one particular characteristic of monopoly, they do not. **Barriers to entry** are the legal, technological, or market forces that discourage or prevent potential competitors from entering a market. Barriers to entry can range from the simple and easily surmountable, such as the cost of renting retail space, to the extremely restrictive. For example, there are a finite number of radio frequencies available for broadcasting. Once the rights to all of them have been purchased, no new competitors can enter the market.

In some cases, barriers to entry may lead to monopoly. In other cases, they may limit competition to a few firms. Barriers may block entry even if the firm or firms currently in the market are earning profits. Thus, in markets with significant barriers to entry, it is *not* true that abnormally high profits will attract new firms, and that this entry of new firms will eventually cause the price to decline so that surviving firms earn only a normal level of profit in the long run.

There are two types of monopoly, based on the types of barriers to entry they exploit. One is **natural monopoly**, where the barriers to entry are something other than legal prohibition. The other is **legal monopoly**, where laws prohibit (or severely limit) competition.

Natural Monopoly

Economies of scale can combine with the size of the market to limit competition. This situation, when economies of scale are large relative to the quantity demanded in the market, is called a natural monopoly. Natural monopolies often arise in industries where the marginal cost of adding an additional customer is very low, once the fixed costs of the overall system are in place. Once the main water pipes are laid through a neighborhood, the marginal cost of providing water service to another home is fairly low. Once electricity lines are installed through a neighborhood, the marginal cost of providing additional electrical service to one more home is very low. It would be costly and duplicative for a second water company to enter the market and invest in a whole second set of main water pipes, or for a second electricity company to enter the market and invest in a whole new set of electrical wires. These industries offer an example where, because of economies of scale, one producer can serve the entire market more efficiently than a number of smaller producers that would need to make duplicate physical capital investments.

A natural monopoly can also arise in smaller local markets for products that are difficult to transport. For example, cement production exhibits economies of scale, and the quantity of cement demanded in a local area may not be much larger than what a single plant can produce. Moreover, the costs of transporting cement over land are high, and so a cement plant in an area without access to water transportation may be a natural monopoly.

Control of a Physical Resource

Another type of natural monopoly occurs when a company has control of a scarce physical resource. In the U.S. economy, one historical example of this pattern occurred when ALCOA—the Aluminum Company of America—controlled most of the supply of bauxite, a key mineral used in making aluminum. Back in the 1930s, when ALCOA controlled most of the bauxite, other firms were simply unable to produce enough aluminum to compete.

As another example, the majority of global diamond production is controlled by DeBeers, a multi-national company that has mining and production operations in South Africa, Botswana, Namibia, and Canada. It

also has exploration activities on four continents, while directing a worldwide distribution network of rough cut diamonds. Though in recent years they have experienced growing competition, their impact on the rough diamond market is still considerable.

Legal Monopoly

For some products, the government erects barriers to entry by prohibiting or limiting competition. Under U.S. law, no organization but the U.S. Postal Service is legally allowed to deliver first-class mail. Many states or cities have laws or regulations that allow households a choice of only one electric company, one water company, and one company to pick up the garbage. Most legal monopolies are considered utilities—products necessary for everyday life—that are socially beneficial to have. As a consequence, the government allows producers to become regulated monopolies, to insure that an appropriate amount of these products is provided to consumers. Additionally, legal monopolies are often subject to economies of scale, so it makes sense to allow only one provider.

Promoting Innovation

Innovation takes time and resources to achieve. Suppose a company invests in research and development and finds the cure for the common cold. In this world of near ubiquitous information, other companies could take the formula, produce the drug, and because they did not incur the costs of research and development (R&D), undercut the price of the company that discovered the drug. Given this possibility, many firms would choose not to invest in research and development, and as a result, the world would have less innovation. To prevent this from happening, the Constitution of the United States specifies in Article I, Section 8: “The Congress shall have Power . . . To Promote the Progress of Science and Useful Arts, by securing for limited Times to Authors and Inventors the Exclusive Right to their Writings and Discoveries.” Congress used this power to create the U.S. Patent and Trademark Office, as well as the U.S. Copyright Office. A **patent** gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time; in the United States, exclusive patent rights last

for 20 years. The idea is to provide limited monopoly power so that innovative firms can recoup their investment in R&D, but then to allow other firms to produce the product more cheaply once the patent expires.

A **trademark** is an identifying symbol or name for a particular good, like Chiquita bananas, Chevrolet cars, or the Nike “swoosh” that appears on shoes and athletic gear. Roughly 1.9 million trademarks are registered with the U.S. government. A firm can renew a trademark over and over again, as long as it remains in active use.

A **copyright**, according to the U.S. Copyright Office, “is a form of protection provided by the laws of the United States for ‘original works of authorship’ including literary, dramatic, musical, architectural, cartographic, choreographic, pantomimic, pictorial, graphic, sculptural, and audiovisual creations.” No one can reproduce, display, or perform a copyrighted work without permission of the author. Copyright protection ordinarily lasts for the life of the author plus 70 years.

Roughly speaking, patent law covers inventions and copyright protects books, songs, and art. But in certain areas, like the invention of new software, it has been unclear whether patent or copyright protection should apply. There is also a body of law known as **trade secrets**. Even if a company does not have a patent on an invention, competing firms are not allowed to steal their secrets. One famous trade secret is the formula for Coca-Cola, which is not protected under copyright or patent law, but is simply kept secret by the company.

Taken together, this combination of patents, trademarks, copyrights, and trade secret law is called **intellectual property**, because it implies ownership over an idea, concept, or image, not a physical piece of property like a house or a car. Countries around the world have enacted laws to protect intellectual property, although the time periods and exact provisions of such laws vary across countries. There are ongoing negotiations, both through the World Intellectual Property Organization (WIPO) and through international treaties, to bring greater harmony to the intellectual property laws of different countries to determine the extent to which patents and copyrights in one country will be respected in other countries.

Government limitations on competition used to be even more common in the United States. For most of the twentieth century, only one phone company—AT&T—was legally allowed to provide local and long distance service. From the 1930s to the 1970s, one set of federal regulations limited which destinations airlines could choose to fly to and what fares they could charge; another set of regulations limited the interest rates that banks could pay to depositors; yet another specified what trucking firms could charge customers.

What products are considered utilities depends, in part, on the available technology. Fifty years ago, local and long distance telephone service was provided over wires. It did not make much sense to have multiple companies building multiple systems of wiring across towns and across the country. AT&T lost its monopoly on long distance service when the technology for providing phone service changed from wires to microwave and satellite transmission, so that multiple firms could use the same transmission mechanism. The same thing happened to local service, especially in recent years, with the growth in cellular phone systems.

The combination of improvements in production technologies and a general sense that the markets could provide services adequately led to a wave of **deregulation**, starting in the late 1970s and continuing into the 1990s. This wave eliminated or reduced government restrictions on the firms that could enter, the prices that could be charged, and the quantities that could be produced in many industries, including telecommunications, airlines, trucking, banking, and electricity.

Around the world, from Europe to Latin America to Africa and Asia, many governments continue to control and limit competition in what those governments perceive to be key industries, including airlines, banks, steel companies, oil companies, and telephone companies.

Intimidating Potential Competition

Businesses have developed a number of schemes for creating barriers to entry by deterring potential competitors from entering the market. One method is known as **predatory pricing**, in which a firm uses the threat of

sharp price cuts to discourage competition. Predatory pricing is a violation of U.S. antitrust law, but it is difficult to prove.

Consider a large airline that provides most of the flights between two particular cities. A new, small start-up airline decides to offer service between these two cities. The large airline immediately slashes prices on this route to the bone, so that the new entrant cannot make any money. After the new entrant has gone out of business, the incumbent firm can raise prices again.

After this pattern is repeated once or twice, potential new entrants may decide that it is not wise to try to compete. Small airlines often accuse larger airlines of predatory pricing: in the early 2000s, for example, ValuJet accused Delta of predatory pricing, Frontier accused United, and Reno Air accused Northwest. In 2015, the Justice Department ruled against American Express and Mastercard for imposing restrictions on retailers who encouraged customers to use lower swipe fees on credit transactions.

In some cases, large advertising budgets can also act as a way of discouraging the competition. If the only way to launch a successful new national cola drink is to spend more than the promotional budgets of Coca-Cola and Pepsi Cola, not too many companies will try. A firmly established brand name can be difficult to dislodge.

Summing Up Barriers to Entry

[\[link\]](#) lists the barriers to entry that have been discussed here. This list is not exhaustive, since firms have proved to be highly creative in inventing business practices that discourage competition. When barriers to entry exist, perfect competition is no longer a reasonable description of how an industry works. When barriers to entry are high enough, monopoly can result.

Barrier to Entry	Government Role?	Example
Natural monopoly	Government often responds with regulation (or ownership)	Water and electric companies
Control of a physical resource	No	DeBeers for diamonds
Legal monopoly	Yes	Post office, past regulation of airlines and trucking
Patent, trademark, and copyright	Yes, through protection of intellectual property	New drugs or software
Intimidating potential competitors	Somewhat	Predatory pricing; well-known brand names

Barriers to Entry

Key Concepts and Summary

Barriers to entry prevent or discourage competitors from entering the market. These barriers include: economies of scale that lead to natural monopoly; control of a physical resource; legal restrictions on competition; patent, trademark and copyright protection; and practices to intimidate the competition like predatory pricing. Intellectual property refers to legally guaranteed ownership of an idea, rather than a physical item. The laws that protect intellectual property include patents, copyrights, trademarks, and trade secrets. A natural monopoly arises when economies of scale persist

over a large enough range of output that if one firm supplies the entire market, no other firm can enter without facing a cost disadvantage.

Self-Check Questions

Exercise:

Problem:

Classify the following as a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or a situation that does not involve a barrier to entry.

- a. A patented invention
- b. A popular but easily copied restaurant recipe
- c. An industry where economies of scale are very small compared to the size of demand in the market
- d. A well-established reputation for slashing prices in response to new entry
- e. A well-respected brand name that has been carefully built up over many years

Solution:

- a. A patent is a government-enforced barrier to entry.
- b. This is not a barrier to entry.
- c. This is not a barrier to entry.
- d. This is a barrier to entry, but it is not government-enforced.
- e. This is a barrier to entry, but it is not directly government enforced.

Exercise:

Problem:

Classify the following as a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or a situation that does not involve a barrier to entry.

- a. A city passes a law on how many licenses it will issue for taxicabs
- b. A city passes a law that all taxicab drivers must pass a driving safety test and have insurance
- c. A well-known trademark
- d. Owning a spring that offers very pure water
- e. An industry where economies of scale are very large compared to the size of demand in the market

Solution:

- a. This is a government-enforced barrier to entry.
- b. This is an example of a government law, but perhaps it is not much of a barrier to entry if most people can pass the safety test and get insurance.
- c. Trademarks are enforced by government, and therefore are a barrier to entry.
- d. This is probably not a barrier to entry, since there are a number of different ways of getting pure water.
- e. This is a barrier to entry, but it is not government-enforced.

Exercise:**Problem:**

Suppose the local electrical utility, a legal monopoly based on economies of scale, was split into four firms of equal size, with the idea that eliminating the monopoly would promote competitive pricing of electricity. What do you anticipate would happen to prices?

Solution:

Because of economies of scale, each firm would produce at a higher average cost than before. (They would each have to build their own power lines.) As a result, they would each have to raise prices to cover their higher costs. The policy would fail.

Exercise:

Problem:

If Congress reduced the period of patent protection from 20 years to 10 years, what would likely happen to the amount of private research and development?

Solution:

Shorter patent protection would make innovation less lucrative, so the amount of research and development would likely decline.

Review Questions

Exercise:

Problem: How is monopoly different from perfect competition?

Exercise:

Problem: What is a barrier to entry? Give some examples.

Exercise:

Problem: What is a natural monopoly?

Exercise:

Problem: What is a legal monopoly?

Exercise:

Problem: What is predatory pricing?

Exercise:

Problem: How is intellectual property different from other property?

Exercise:

Problem:

By what legal mechanisms is intellectual property protected?

Exercise:

Problem: In what sense is a natural monopoly “natural”?

Critical Thinking Questions

Exercise:

Problem:

ALCOA does not have the monopoly power it once had. How do you suppose their barriers to entry were weakened?

Exercise:

Problem:

Why are generic pharmaceuticals significantly cheaper than name brand ones?

Exercise:

Problem:

For many years, the Justice Department has tried to break up large firms like IBM, Microsoft, and most recently Google, on the grounds that their large market share made them essentially monopolies. In a global market, where U.S. firms compete with firms from other countries, would this policy make the same sense as it might in a purely domestic context?

Exercise:**Problem:**

Intellectual property laws are intended to promote innovation, but some economists, such as Milton Friedman, have argued that such laws are not desirable. In the United States, there is no intellectual property protection for food recipes or for fashion designs. Considering the state of these two industries, and bearing in mind the discussion of the inefficiency of monopolies, can you think of any reasons why intellectual property laws might hinder innovation in some cases?

Problems**Exercise:****Problem:**

Return to [\[link\]](#). Suppose P_0 is \$10 and P_1 is \$11. Suppose a new firm with the same LRAC curve as the incumbent tries to break into the market by selling 4,000 units of output. Estimate from the graph what the new firm's average cost of producing output would be. If the incumbent continues to produce 6,000 units, how much output would be supplied to the market by the two firms? Estimate what would happen to the market price as a result of the supply of both the incumbent firm and the new entrant. Approximately how much profit would each firm earn?

Glossary

barriers to entry

the legal, technological, or market forces that may discourage or prevent potential competitors from entering a market

copyright

a form of legal protection to prevent copying, for commercial purposes, original works of authorship, including books and music

deregulation

removing government controls over setting prices and quantities in certain industries

intellectual property

the body of law including patents, trademarks, copyrights, and trade secret law that protect the right of inventors to produce and sell their inventions

legal monopoly

legal prohibitions against competition, such as regulated monopolies and intellectual property protection

monopoly

a situation in which one firm produces all of the output in a market

natural monopoly

economic conditions in the industry, for example, economies of scale or control of a critical resource, that limit effective competition

patent

a government rule that gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time

predatory pricing

when an existing firm uses sharp but temporary price cuts to discourage new competition

trade secrets

methods of production kept secret by the producing firm

trademark

an identifying symbol or name for a particular good and can only be used by the firm that registered that trademark

How a Profit-Maximizing Monopoly Chooses Output and Price

By the end of this section, you will be able to:

- Explain the perceived demand curve for a perfect competitor and a monopoly
- Analyze a demand curve for a monopoly and determine the output that maximizes profit and revenue
- Calculate marginal revenue and marginal cost
- Explain allocative efficiency as it pertains to the efficiency of a monopoly

Consider a monopoly firm, comfortably surrounded by barriers to entry so that it need not fear competition from other producers. How will this monopoly choose its profit-maximizing quantity of output, and what price will it charge? Profits for the monopolist, like any firm, will be equal to total revenues minus total costs. The pattern of costs for the monopoly can be analyzed within the same framework as the costs of a perfectly competitive firm—that is, by using total cost, fixed cost, variable cost, marginal cost, average cost, and average variable cost. However, because a monopoly faces no competition, its situation and its decision process will differ from that of a perfectly competitive firm. (The Clear it Up feature discusses how hard it is sometimes to define “market” in a monopoly situation.)

Demand Curves Facing a Monopoly

While a monopolist can charge *any* price for its product, that price is nonetheless constrained by demand for the firm’s product. No monopolist, even one that is thoroughly protected by high barriers to entry, can require consumers to purchase its product. Because the monopolist is the only firm in the market, its demand curve is the same as the market demand curve, which is, unlike that for a perfectly competitive firm, downward-sloping

Note:

What defines the market?

A monopoly is a firm that sells all or nearly all of the goods and services in a given market. But what defines the “market”?

In a famous 1947 case, the federal government accused the DuPont company of having a monopoly in the cellophane market, pointing out that DuPont produced 75% of the cellophane in the United States. DuPont countered that even though it had a 75% market share in cellophane, it had less than a 20% share of the “flexible packaging materials,” which includes all other moisture-proof papers, films, and foils. In 1956, after years of legal appeals, the U.S. Supreme Court held that the broader market definition was more appropriate, and the case against DuPont was dismissed.

Questions over how to define the market continue today. True, Microsoft in the 1990s had a dominant share of the software for computer operating systems, but in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014. The Greyhound bus company may have a near-monopoly on the market for intercity bus transportation, but it is only a small share of the market for intercity transportation if that market includes private cars, airplanes, and railroad service. DeBeers has a monopoly in diamonds, but it is a much smaller share of the total market for precious gemstones and an even smaller share of the total market for jewelry. A small town in the country may have only one gas station: is this gas station a “monopoly,” or does it compete with gas stations that might be five, 10, or 50 miles away?

In general, if a firm produces a product without close substitutes, then the firm can be considered a monopoly producer in a single market. But if buyers have a range of similar—even if not identical—options available from other firms, then the firm is not a monopoly. Still, arguments over whether substitutes are close or not close can be controversial.

The Inefficiency of Monopoly

Most people criticize monopolies because they charge too high a price, but what economists object to is that monopolies do not supply enough output

to be allocatively efficient. To understand why a monopoly is inefficient, it is useful to compare it with the benchmark model of perfect competition.

Allocative efficiency is a social concept. It refers to producing the optimal quantity of some output, the quantity where the marginal benefit to society of one more unit just equals the marginal cost. The rule of profit maximization in a world of perfect competition was for each firm to produce the quantity of output where $P = MC$, where the price (P) is a measure of how much buyers value the good and the marginal cost (MC) is a measure of what marginal units cost society to produce. Following this rule assures allocative efficiency. If $P > MC$, then the marginal benefit to society (as measured by P) is greater than the marginal cost to society of producing additional units, and a greater quantity should be produced. But in the case of monopoly, price is always greater than marginal cost at the profit-maximizing level of output, as can be seen by looking back at [\[link\]](#). Thus, consumers will suffer from a monopoly because a lower quantity will be sold in the market, at a higher price, than would have been the case in a perfectly competitive market.

The problem of inefficiency for monopolies often runs even deeper than these issues, and also involves incentives for efficiency over longer periods of time. There are counterbalancing incentives here. On one side, firms may strive for new inventions and new intellectual property because they want to become monopolies and earn high profits—at least for a few years until the competition catches up. In this way, monopolies may come to exist because of competitive pressures on firms. However, once a barrier to entry is in place, a monopoly that does not need to fear competition can just produce the same old products in the same old way—while still ringing up a healthy rate of profit. John Hicks, who won the Nobel Prize for economics in 1972, wrote in 1935: “The best of all monopoly profits is a quiet life.” He did not mean the comment in a complimentary way. He meant that monopolies may bank their profits and slack off on trying to please their customers.

When AT&T provided all of the local and long-distance phone service in the United States, along with manufacturing most of the phone equipment, the payment plans and types of phones did not change much. The old joke was that you could have any color phone you wanted, as long as it was

black. But in 1982, AT&T was split up by government litigation into a number of local phone companies, a long-distance phone company, and a phone equipment manufacturer. An explosion of innovation followed. Services like call waiting, caller ID, three-way calling, voice mail though the phone company, mobile phones, and wireless connections to the Internet all became available. A wide range of payment plans was offered, as well. It was no longer true that all phones were black; instead, phones came in a wide variety of shapes and colors. The end of the telephone monopoly brought lower prices, a greater quantity of services, and also a wave of innovation aimed at attracting and pleasing customers.

Note:**The Rest is History**

In the opening case, the East India Company and the Confederate States were presented as a monopoly or near monopoly provider of a good.

Nearly every American schoolchild knows the result of the ‘unwelcome visit’ the ‘Mohawks’ bestowed upon Boston Harbor’s tea-bearing ships—the Boston Tea Party. Regarding the cotton industry, we also know Great Britain remained neutral during the Civil War, taking neither side during the conflict.

Did the monopoly nature of these business have unintended and historical consequences? Might the American Revolution have been deterred, if the East India Company had sailed the tea-bearing ships back to England? Might the southern states have made different decisions had they not been so confident “King Cotton” would force diplomatic recognition of the Confederate States of America? Of course, it is not possible to definitively answer these questions; after all we cannot roll back the clock and try a different scenario. We can, however, consider the monopoly nature of these businesses and the roles they played and hypothesize about what might have occurred under different circumstances.

Perhaps if there had been legal free tea trade, the colonists would have seen things differently; there was smuggled Dutch tea in the colonial market. If the colonists had been able to freely purchase Dutch tea, they would have paid lower prices and avoided the tax.

What about the cotton monopoly? With one in five jobs in Great Britain depending on Southern cotton and the Confederate States nearly the sole provider of that cotton, why did Great Britain remain neutral during the Civil War? At the beginning of the war, Britain simply drew down massive stores of cotton. These stockpiles lasted until near the end of 1862. Why did Britain not recognize the Confederacy at that point? Two reasons: The Emancipation Proclamation and new sources of cotton. Having outlawed slavery throughout the United Kingdom in 1833, it was politically impossible for Great Britain, empty cotton warehouses or not, to recognize, diplomatically, the Confederate States. In addition, during the two years it took to draw down the stockpiles, Britain expanded cotton imports from India, Egypt, and Brazil.

Monopoly sellers often see no threats to their superior marketplace position. In these examples did the power of the monopoly blind the decision makers to other possibilities? Perhaps. But, as they say, the rest is history.

Key Concepts and Summary

A monopolist is not a price taker, because when it decides what quantity to produce, it also determines the market price. For a monopolist, total revenue is relatively low at low quantities of output, because not much is being sold. Total revenue is also relatively low at very high quantities of output, because a very high quantity will sell only at a low price. Thus, total revenue for a monopolist will start low, rise, and then decline. The marginal revenue for a monopolist from selling additional units will decline. Each additional unit sold by a monopolist will push down the overall market price, and as more units are sold, this lower price applies to more and more units.

The monopolist will select the profit-maximizing level of output where $MR = MC$, and then charge the price for that quantity of output as determined by the market demand curve. If that price is above average cost, the monopolist earns positive profits.

Monopolists are not productively efficient, because they do not produce at the minimum of the average cost curve. Monopolists are not allocatively efficient, because they do not produce at the quantity where $P = MC$. As a result, monopolists produce less, at a higher average cost, and charge a higher price than would a combination of firms in a perfectly competitive industry. Monopolists also may lack incentives for innovation, because they need not fear entry.

Self-Check Questions

Exercise:

Problem:

Suppose demand for a monopoly's product falls so that its profit-maximizing price is below average variable cost. How much output should the firm supply? *Hint:* Draw the graph.

Solution:

If price falls below AVC, the firm will not be able to earn enough revenues even to cover its variable costs. In such a case, it will suffer a smaller loss if it shuts down and produces no output. By contrast, if it stayed in operation and produced the level of output where $MR = MC$, it would lose all of its fixed costs plus some variable costs. If it shuts down, it only loses its fixed costs.

Exercise:

Problem:

Imagine a monopolist could charge a different price to every customer based on how much he or she were willing to pay. How would this affect monopoly profits?

Solution:

This scenario is called "perfect price discrimination." The result would be that the monopolist would produce more output, the same amount

in fact as would be produced by a perfectly competitive industry. However, there would be no consumer surplus since each buyer is paying exactly what they think the product is worth. Therefore, the monopolist would be earning the maximum possible profits.

Review Questions

Exercise:

Problem:

How is the demand curve perceived by a perfectly competitive firm different from the demand curve perceived by a monopolist?

Exercise:

Problem:

How does the demand curve perceived by a monopolist compare with the market demand curve?

Exercise:

Problem: Is a monopolist a price taker? Explain briefly.

Exercise:

Problem:

What is the usual shape of a total revenue curve for a monopolist? Why?

Exercise:

Problem:

What is the usual shape of a marginal revenue curve for a monopolist? Why?

Exercise:

Problem:

How can a monopolist identify the profit-maximizing level of output if it knows its total revenue and total cost curves?

Exercise:

Problem:

How can a monopolist identify the profit-maximizing level of output if it knows its marginal revenue and marginal costs?

Exercise:

Problem:

When a monopolist identifies its profit-maximizing quantity of output, how does it decide what price to charge?

Exercise:

Problem: Is a monopolist allocatively efficient? Why or why not?

Exercise:

Problem:

How does the quantity produced and price charged by a monopolist compare to that of a perfectly competitive firm?

Critical Thinking Questions

Exercise:

Problem:

Imagine that you are managing a small firm and thinking about entering the market of a monopolist. The monopolist is currently charging a high price, and you have calculated that you can make a nice profit charging 10% less than the monopolist. Before you go ahead and challenge the monopolist, what possibility should you consider for how the monopolist might react?

Exercise:**Problem:**

If a monopoly firm is earning profits, how much would you expect these profits to be diminished by entry in the long run?

Problems**Exercise:****Problem:**

Draw the demand curve, marginal revenue, and marginal cost curves from [\[link\]](#), and identify the quantity of output the monopoly wishes to supply and the price it will charge. Suppose demand for the monopoly's product increases dramatically. Draw the new demand curve. What happens to the marginal revenue as a result of the increase in demand? What happens to the marginal cost curve? Identify the new profit-maximizing quantity and price. Does the answer make sense to you?

Exercise:

Problem:

Draw a monopolist's demand curve, marginal revenue, and marginal cost curves. Identify the monopolist's profit-maximizing output level. Now, think about a slightly higher level of output (say $Q_0 + 1$). According to the graph, is there any consumer willing to pay more than the marginal cost of that new level of output? If so, what does this mean?

References

Aboukhadijeh, Feross. "Chapter 20: Girding for War - The North and the South, 1861-1865." *StudyNotes, Inc.* Accessed July 7, 2013. <http://www.apstudynotes.org/us-history/outlines/chapter-20-girding-for-war-the-north-and-the-south-1861-1865/>.

British Parliament. "(28 August 1833). Slavery Abolition Act 1833; Section LXIV." Accessed July 2013. http://www.pdavis.nl/Legis_07.htm.

Dattel, E. (nd). "Cotton and the Civil War." *Mississippi Historical Society*. Accessed July 2013. <http://mshistorynow.mdah.state.ms.us/articles/291/cotton-and-the-civil-war>.

Gartner. 2015. "Gartner Says Tablet Sales Continue to Be Slow in 2015." Accessed March 12, 2015. <http://www.gartner.com/newsroom/id/2954317>.

Grogan, David. 2015. "Federal Judge Finds AmEx's Anti-Steering Rule Violates Antitrust Law." American Booksellers Association. Accessed March 12, 2015. <http://www.bookweb.org/news/federal-judge-finds-amex%E2%80%99s-anti-steering-rule-violates-antitrust-law>.

Massachusetts Historical Society. "The Coming of the American Revolution 1764-1776: The Boston Tea Party." Retrieved from <http://www.masshist.org/revolution/teaparty.php>.

Massachusetts Historical Society. "Whereas our Nation." *The Massachusetts Gazette*, p. 2. Accessed July 2013

http://www.masshist.org/revolution/image-viewer.php?old=1&item_id=457&img_step=1&nmask=1&mode=large.

Pelegrin, William. 2015. "Judge Overrules Antitrust Case Against Google , Says Setting Default Search Engines Is Fair." Digital Trends. Accessed March 12, 2015. <http://www.digitaltrends.com/mobile/judge-tosses-out-google-antitrust-lawsuit/>.

Glossary

allocative efficiency

producing the optimal quantity of some output; the quantity where the marginal benefit to society of one more unit just equals the marginal cost

marginal profit

profit of one more unit of output, computed as marginal revenue minus marginal cost

Introduction to Monopolistic Competition and Oligopoly

class="introduction"

Competing Brands?

The laundry
detergent
market is
one that is
characterize
d neither as
perfect
competition
nor
monopoly.

(Credit:
modification
of work by
Pixel
Drip/Flickr
Creative
Commons)

**Note:****The Temptation to Defy the Law**

Laundry detergent and bags of ice—products of industries that seem pretty mundane, maybe even boring. Hardly! Both have been the center of clandestine meetings and secret deals worthy of a spy novel. In France, between 1997 and 2004, the top four laundry detergent producers (Proctor & Gamble, Henkel, Unilever, and Colgate-Palmolive) controlled about 90 percent of the French soap market. Officials from the soap firms were meeting secretly, in out-of-the-way, small cafés around Paris. Their goals: Stamp out competition and set prices.

Around the same time, the top five Midwest ice makers (Home City Ice, Lang Ice, Tinley Ice, Sisler's Dairy, and Products of Ohio) had similar goals in mind when they secretly agreed to divide up the bagged ice market.

If both groups could meet their goals, it would enable each to act as though they were a single firm—in essence, a monopoly—and enjoy monopoly-size profits. The problem? In many parts of the world, including the

European Union and the United States, it is illegal for firms to divide up markets and set prices collaboratively.

These two cases provide examples of markets that are characterized neither as perfect competition nor monopoly. Instead, these firms are competing in market structures that lie between the extremes of monopoly and perfect competition. How do they behave? Why do they exist? We will revisit this case later, to find out what happened.

Note:

Introduction to Monopolistic Competition and Oligopoly

In this chapter, you will learn about:

- Monopolistic Competition
- Oligopoly

Perfect competition and monopoly are at opposite ends of the competition spectrum. A perfectly competitive market has many firms selling identical products, who all act as price takers in the face of the competition. If you recall, price takers are firms that have no market power. They simply have to take the market price as given.

Monopoly arises when a single firm sells a product for which there are no close substitutes. Microsoft, for instance, has been considered a monopoly because of its domination of the operating systems market.

What about the vast majority of real world firms and organizations that fall between these extremes, firms that could be described as **imperfectly competitive**? What determines their behavior? They have more influence over the price they charge than perfectly competitive firms, but not as much as a monopoly would. What will they do?

One type of imperfectly competitive market is called **monopolistic competition**. Monopolistically competitive markets feature a large number

of competing firms, but the products that they sell are not identical. Consider, as an example, the Mall of America in Minnesota, the largest shopping mall in the United States. In 2010, the Mall of America had 24 stores that sold women's "ready-to-wear" clothing (like Ann Taylor and Urban Outfitters), another 50 stores that sold clothing for both men and women (like Banana Republic, J. Crew, and Nordstrom's), plus 14 more stores that sold women's specialty clothing (like Motherhood Maternity and Victoria's Secret). Most of the markets that consumers encounter at the retail level are monopolistically competitive.

The other type of imperfectly competitive market is **oligopoly**. Oligopolistic markets are those dominated by a small number of firms. Commercial aircraft provides a good example: Boeing and Airbus each produce slightly less than 50% of the large commercial aircraft in the world. Another example is the U.S. soft drink industry, which is dominated by Coca-Cola and Pepsi. Oligopolies are characterized by high barriers to entry with firms choosing output, pricing, and other decisions strategically based on the decisions of the other firms in the market. In this chapter, we first explore how monopolistically competitive firms will choose their profit-maximizing level of output. We will then discuss oligopolistic firms, which face two conflicting temptations: to collaborate as if they were a single monopoly, or to individually compete to gain profits by expanding output levels and cutting prices. Oligopolistic markets and firms can also take on elements of monopoly and of perfect competition.

Monopolistic Competition

By the end of this section, you will be able to:

- Explain the significance of differentiated products
- Describe how a monopolistic competitor chooses price and quantity
- Discuss entry, exit, and efficiency as they pertain to monopolistic competition
- Analyze how advertising can impact monopolistic competition

Monopolistic competition involves many firms competing against each other, but selling products that are distinctive in some way. Examples include stores that sell different styles of clothing; restaurants or grocery stores that sell different kinds of food; and even products like golf balls or beer that may be at least somewhat similar but differ in public perception because of advertising and brand names. There are over 600,000 restaurants in the United States. When products are distinctive, each firm has a mini-monopoly on its particular style or flavor or brand name. However, firms producing such products must also compete with other styles and flavors and brand names. The term “monopolistic competition” captures this mixture of mini-monopoly and tough competition, and the following Clear It Up feature introduces its derivation.

Note:

Who invented the theory of imperfect competition?

The theory of imperfect competition was developed by two economists independently but simultaneously in 1933. The first was Edward Chamberlin of Harvard University who published *The Economics of Monopolistic Competition*. The second was Joan Robinson of Cambridge University who published *The Economics of Imperfect Competition*. Robinson subsequently became interested in macroeconomics where she became a prominent Keynesian, and later a post-Keynesian economist. (See the [Welcome to Economics!](#) and [The Keynesian Perspective](#) chapters for more on Keynes.)

Differentiated Products

A firm can try to make its products different from those of its competitors in several ways: physical aspects of the product, location from which the product is sold, intangible aspects of the product, and perceptions of the product. Products that are distinctive in one of these ways are called **differentiated products**.

Physical aspects of a product include all the phrases you hear in advertisements: unbreakable bottle, nonstick surface, freezer-to-microwave, non-shrink, extra spicy, newly redesigned for your comfort. The location of a firm can also create a difference between producers. For example, a gas station located at a heavily traveled intersection can probably sell more gas, because more cars drive by that corner. A supplier to an automobile manufacturer may find that it is an advantage to locate close to the car factory.

Intangible aspects can differentiate a product, too. Some intangible aspects may be promises like a guarantee of satisfaction or money back, a reputation for high quality, services like free delivery, or offering a loan to purchase the product. Finally, product differentiation may occur in the minds of buyers. For example, many people could not tell the difference in taste between common varieties of beer or cigarettes if they were blindfolded but, because of past habits and advertising, they have strong preferences for certain brands. Advertising can play a role in shaping these intangible preferences.

The concept of differentiated products is closely related to the degree of variety that is available. If everyone in the economy wore only blue jeans, ate only white bread, and drank only tap water, then the markets for clothing, food, and drink would be much closer to perfectly competitive. The variety of styles, flavors, locations, and characteristics creates product differentiation and monopolistic competition.

Note:

Are golf balls really differentiated products?

Monopolistic competition refers to an industry that has more than a few firms, each offering a product which, from the consumer's perspective, is different from its competitors. The U.S. Golf Association runs a laboratory that tests 20,000 golf balls a year. There are strict rules for what makes a golf ball legal. The weight of a golf ball cannot exceed 1.620 ounces and its diameter cannot be less than 1.680 inches (which is a weight of 45.93 grams and a diameter of 42.67 millimeters, in case you were wondering). The balls are also tested by being hit at different speeds. For example, the distance test involves having a mechanical golfer hit the ball with a titanium driver and a swing speed of 120 miles per hour. As the testing center explains: "The USGA system then uses an array of sensors that accurately measure the flight of a golf ball during a short, indoor trajectory from a ball launcher. From this flight data, a computer calculates the lift and drag forces that are generated by the speed, spin, and dimple pattern of the ball. ... The distance limit is 317 yards."

Over 1800 golf balls made by more than 100 companies meet the USGA standards. The balls do differ in various ways, like the pattern of dimples on the ball, the types of plastic used on the cover and in the cores, and so on. Since all balls need to conform to the USGA tests, they are much more alike than different. In other words, golf ball manufacturers are monopolistically competitive.

However, retail sales of golf balls are about \$500 million per year, which means that a lot of large companies have a powerful incentive to persuade players that golf balls are highly differentiated and that it makes a huge difference which one you choose. Sure, Tiger Woods can tell the difference. For the average duffer (golf-speak for a "mediocre player") who plays a few times a summer—and who loses a lot of golf balls to the woods and lake and needs to buy new ones—most golf balls are pretty much indistinguishable.

Monopolistic Competitors and Entry

If one monopolistic competitor earns positive economic profits, other firms will be tempted to enter the market. A gas station with a great location must worry that other gas stations might open across the street or down the road

—and perhaps the new gas stations will sell coffee or have a carwash or some other attraction to lure customers. A successful restaurant with a unique barbecue sauce must be concerned that other restaurants will try to copy the sauce or offer their own unique recipes. A laundry detergent with a great reputation for quality must be concerned that other competitors may seek to build their own reputations.

Monopolistic competitors can make an economic profit or loss in the short run, but in the long run, entry and exit will drive these firms toward a zero economic profit outcome. However, the zero economic profit outcome in monopolistic competition looks different from the zero economic profit outcome in perfect competition in several ways relating both to efficiency and to variety in the market.

Monopolistic Competition and Efficiency

The long-term result of entry and exit in a perfectly competitive market is that all firms end up selling at the price level determined by the lowest point on the average cost curve. This outcome is why perfect competition displays productive efficiency: goods are being produced at the lowest possible average cost. However, in monopolistic competition, the end result of entry and exit is that firms end up with a price that lies on the downward-sloping portion of the average cost curve, not at the very bottom of the AC curve. Thus, monopolistic competition will not be productively efficient.

In a perfectly competitive market, each firm produces at a quantity where price is set equal to marginal cost, both in the short run and in the long run. This outcome is why perfect competition displays allocative efficiency: the social benefits of additional production, as measured by the marginal benefit, which is the same as the price, equal the marginal costs to society of that production. In a monopolistically competitive market, the rule for maximizing profit is to set $MR = MC$ —and price is higher than marginal revenue, not equal to it because the demand curve is downward sloping. When $P > MC$, which is the outcome in a monopolistically competitive market, the benefits to society of providing additional quantity, as measured by the price that people are willing to pay, exceed the marginal costs to society of producing those units. A monopolistically competitive firm does

not produce more, which means that society loses the net benefit of those extra units. This is the same argument we made about monopoly, but in this case to a lesser degree. Thus, a monopolistically competitive industry will produce a lower quantity of a good and charge a higher price for it than would a perfectly competitive industry. See the following Clear It Up feature for more detail on the impact of demand shifts.

A monopolistically competitive industry does not display productive and allocative efficiency in either the short run, when firms are making economic profits and losses, nor in the long run, when firms are earning zero profits.

The Benefits of Variety and Product Differentiation

Even though monopolistic competition does not provide productive efficiency or allocative efficiency, it does have benefits of its own. Product differentiation is based on variety and innovation. Many people would prefer to live in an economy with many kinds of clothes, foods, and car styles; not in a world of perfect competition where everyone will always wear blue jeans and white shirts, eat only spaghetti with plain red sauce, and drive an identical model of car. Many people would prefer to live in an economy where firms are struggling to figure out ways of attracting customers by methods like friendlier service, free delivery, guarantees of quality, variations on existing products, and a better shopping experience.

Economists have struggled, with only partial success, to address the question of whether a market-oriented economy produces the optimal amount of variety. Critics of market-oriented economies argue that society does not really need dozens of different athletic shoes or breakfast cereals or automobiles. They argue that much of the cost of creating such a high degree of product differentiation, and then of advertising and marketing this differentiation, is socially wasteful—that is, most people would be just as happy with a smaller range of differentiated products produced and sold at a lower price. Defenders of a market-oriented economy respond that if people do not want to buy differentiated products or highly advertised brand names, no one is forcing them to do so. Moreover, they argue that consumers benefit substantially when firms seek short-term profits by

providing differentiated products. This controversy may never be fully resolved, in part because deciding on the optimal amount of variety is very difficult, and in part because the two sides often place different values on what variety means for consumers. Read the following Clear It Up feature for a discussion on the role that advertising plays in monopolistic competition.

Note:

How does advertising impact monopolistic competition?

The U.S. economy spent about \$180.12 billion on advertising in 2014, according to eMarketer.com. Roughly one third of this was television advertising, and another third was divided roughly equally between Internet, newspapers, and radio. The remaining third was divided up between direct mail, magazines, telephone directory yellow pages, and billboards. Mobile devices are increasing the opportunities for advertisers. Advertising is all about explaining to people, or making people believe, that the products of one firm are differentiated from the products of another firm. In the framework of monopolistic competition, there are two ways to conceive of how advertising works: either advertising causes a firm's perceived demand curve to become more inelastic (that is, it causes the perceived demand curve to become steeper); or advertising causes demand for the firm's product to increase (that is, it causes the firm's perceived demand curve to shift to the right). In either case, a successful advertising campaign may allow a firm to sell either a greater quantity or to charge a higher price, or both, and thus increase its profits.

However, economists and business owners have also long suspected that much of the advertising may only offset other advertising. Economist A. C. Pigou wrote the following back in 1920 in his book, *The Economics of Welfare*:

"It may happen that expenditures on advertisement made by competing monopolists [that is, what we now call monopolistic competitors] will simply neutralise one another, and leave the industrial position exactly as it would have been if neither had expended anything. For, clearly, if each of two rivals makes equal efforts to attract the favour of the public away from

the other, the total result is the same as it would have been if neither had made any effort at all."

Key Concepts and Summary

Monopolistic competition refers to a market where many firms sell differentiated products. Differentiated products can arise from characteristics of the good or service, location from which the product is sold, intangible aspects of the product, and perceptions of the product.

The perceived demand curve for a monopolistically competitive firm is downward-sloping, which shows that it is a price maker and chooses a combination of price and quantity. However, the perceived demand curve for a monopolistic competitor is more elastic than the perceived demand curve for a monopolist, because the monopolistic competitor has direct competition, unlike the pure monopolist. A profit-maximizing monopolistic competitor will seek out the quantity where marginal revenue is equal to marginal cost. The monopolistic competitor will produce that level of output and charge the price that is indicated by the firm's demand curve.

If the firms in a monopolistically competitive industry are earning economic profits, the industry will attract entry until profits are driven down to zero in the long run. If the firms in a monopolistically competitive industry are suffering economic losses, then the industry will experience exit of firms until economic profits are driven up to zero in the long run.

A monopolistically competitive firm is not productively efficient because it does not produce at the minimum of its average cost curve. A monopolistically competitive firm is not allocatively efficient because it does not produce where $P = MC$, but instead produces where $P > MC$. Thus, a monopolistically competitive firm will tend to produce a lower quantity at a higher cost and to charge a higher price than a perfectly competitive firm.

Monopolistically competitive industries do offer benefits to consumers in the form of greater variety and incentives for improved products and

services. There is some controversy over whether a market-oriented economy generates too much variety.

Self-Check Questions

Exercise:

Problem:

Suppose that, due to a successful advertising campaign, a monopolistic competitor experiences an increase in demand for its product. How will that affect the price it charges and the quantity it supplies?

Solution:

An increase in demand will manifest itself as a rightward shift in the demand curve, and a rightward shift in marginal revenue. The shift in marginal revenue will cause a movement up the marginal cost curve to the new intersection between MR and MC at a higher level of output. The new price can be read by drawing a line up from the new output level to the new demand curve, and then over to the vertical axis. The new price should be higher. The increase in quantity will cause a movement along the average cost curve to a possibly higher level of average cost. The price, though, will increase more, causing an increase in total profits.

Exercise:

Problem:

Continuing with the scenario outlined in question 1, in the long run, the positive economic profits earned by the monopolistic competitor will attract a response either from existing firms in the industry or firms outside. As those firms capture the original firm's profit, what will happen to the original firm's profit-maximizing price and output levels?

Solution:

As long as the original firm is earning positive economic profits, other firms will respond in ways that take away the original firm's profits. This will manifest itself as a decrease in demand for the original firm's product, a decrease in the firm's profit-maximizing price and a decrease in the firm's profit-maximizing level of output, essentially unwinding the process described in the answer to question 1. In the long-run equilibrium, all firms in monopolistically competitive markets will earn zero economic profits.

Review Questions

Exercise:

Problem:

What is the relationship between product differentiation and monopolistic competition?

Exercise:

Problem:

How is the perceived demand curve for a monopolistically competitive firm different from the perceived demand curve for a monopoly or a perfectly competitive firm?

Exercise:

Problem:

How does a monopolistic competitor choose its profit-maximizing quantity of output and price?

Exercise:

Problem:

How can a monopolistic competitor tell whether the price it is charging will cause the firm to earn profits or experience losses?

Exercise:

Problem:

If the firms in a monopolistically competitive market are earning economic profits or losses in the short run, would you expect them to continue doing so in the long run? Why?

Exercise:**Problem:**

Is a monopolistically competitive firm productively efficient? Is it allocatively efficient? Why or why not?

Critical Thinking Questions**Exercise:****Problem:**

Aside from advertising, how can monopolistically competitive firms increase demand for their products?

Exercise:**Problem:**

Make a case for why monopolistically competitive industries never reach long-run equilibrium.

Exercise:**Problem:**

Would you rather have efficiency or variety? That is, one opportunity cost of the variety of products we have is that each product costs more per unit than if there were only one kind of product of a given type, like shoes. Perhaps a better question is, “What is the right amount of variety? Can there be too many varieties of shoes, for example?”

Problems

Exercise:

Problem:

Andrea's Day Spa began to offer a relaxing aromatherapy treatment. The firm asks you how much to charge to maximize profits. The demand curve for the treatments is given by the first two columns in [\[link\]](#); its total costs are given in the third column. For each level of output, calculate total revenue, marginal revenue, average cost, and marginal cost. What is the profit-maximizing level of output for the treatments and how much will the firm earn in profits?

Price	Quantity	TC
\$25.00	0	\$130
\$24.00	10	\$275
\$23.00	20	\$435
\$22.50	30	\$610
\$22.00	40	\$800
\$21.60	50	\$1,005
\$21.20	60	\$1,225

References

Kantar Media. “Our Insights: Infographic—U.S. Advertising Year End Trends Report 2012.” Accessed October 17, 2013.
<http://kantarmedia.us/insight-center/reports/infographic-us-advertising-year-end-trends-report-2012>.

Statistica.com. 2015. “Number of Restaurants in the United States from 2011 to 2014.” Accessed March 27, 2015.
<http://www.statista.com/statistics/244616/number-of-qsr-fsr-chain-independent-restaurants-in-the-us/>.

Glossary

differentiated product

a product that is perceived by consumers as distinctive in some way

imperfectly competitive

firms and organizations that fall between the extremes of monopoly and perfect competition

monopolistic competition

many firms competing to sell similar but differentiated products

oligopoly

when a few large firms have all or most of the sales in an industry

Oligopoly

By the end of this section, you will be able to:

- Explain why and how oligopolies exist
- Contrast collusion and competition
- Interpret and analyze the prisoner's dilemma diagram
- Evaluate the tradeoffs of imperfect competition

Many purchases that individuals make at the retail level are produced in markets that are neither perfectly competitive, monopolies, nor monopolistically competitive. Rather, they are oligopolies. Oligopoly arises when a small number of large firms have all or most of the sales in an industry. Examples of oligopoly abound and include the auto industry, cable television, and commercial air travel. Oligopolistic firms are like cats in a bag. They can either scratch each other to pieces or cuddle up and get comfortable with one another. If oligopolists compete hard, they may end up acting very much like perfect competitors, driving down costs and leading to zero profits for all. If oligopolists collude with each other, they may effectively act like a monopoly and succeed in pushing up prices and earning consistently high levels of profit. Oligopolies are typically characterized by mutual interdependence where various decisions such as output, price, advertising, and so on, depend on the decisions of the other firm(s). Analyzing the choices of oligopolistic firms about pricing and quantity produced involves considering the pros and cons of competition versus collusion at a given point in time.

Why Do Oligopolies Exist?

A combination of the barriers to entry that create monopolies and the product differentiation that characterizes monopolistic competition can create the setting for an oligopoly. For example, when a government grants a patent for an invention to one firm, it may create a monopoly. When the government grants patents to, for example, three different pharmaceutical companies that each has its own drug for reducing high blood pressure, those three firms may become an oligopoly.

Similarly, a natural monopoly will arise when the quantity demanded in a market is only large enough for a single firm to operate at the minimum of the long-run average cost curve. In such a setting, the market has room for only one firm, because no smaller firm can operate at a low enough average cost to compete, and no larger firm could sell what it produced given the quantity demanded in the market.

Quantity demanded in the market may also be two or three times the quantity needed to produce at the minimum of the average cost curve—which means that the market would have room for only two or three oligopoly firms (and they need not produce differentiated products). Again, smaller firms would have higher average costs and be unable to compete, while additional large firms would produce such a high quantity that they would not be able to sell it at a profitable price. This combination of economies of scale and market demand creates the barrier to entry, which led to the Boeing-Airbus oligopoly for large passenger aircraft.

The product differentiation at the heart of monopolistic competition can also play a role in creating oligopoly. For example, firms may need to reach a certain minimum size before they are able to spend enough on advertising and marketing to create a recognizable brand name. The problem in competing with, say, Coca-Cola or Pepsi is not that producing fizzy drinks is technologically difficult, but rather that creating a brand name and marketing effort to equal Coke or Pepsi is an enormous task.

Collusion or Competition?

When oligopoly firms in a certain market decide what quantity to produce and what price to charge, they face a temptation to act as if they were a monopoly. By acting together, oligopolistic firms can hold down industry output, charge a higher price, and divide up the profit among themselves. When firms act together in this way to reduce output and keep prices high, it is called **collusion**. A group of firms that have a formal agreement to collude to produce the monopoly output and sell at the monopoly price is called a **cartel**. See the following Clear It Up feature for a more in-depth analysis of the difference between the two.

Note:

Collusion versus cartels: How can I tell which is which?

In the United States, as well as many other countries, it is illegal for firms to collude since collusion is anti-competitive behavior, which is a violation of antitrust law. Both the Antitrust Division of the Justice Department and the Federal Trade Commission have responsibilities for preventing collusion in the United States.

The problem of enforcement is finding hard evidence of collusion. Cartels are formal agreements to collude. Because cartel agreements provide evidence of collusion, they are rare in the United States. Instead, most collusion is tacit, where firms implicitly reach an understanding that competition is bad for profits.

The desire of businesses to avoid competing so that they can instead raise the prices that they charge and earn higher profits has been well understood by economists. Adam Smith wrote in *Wealth of Nations* in 1776: “People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices.”

Even when oligopolists recognize that they would benefit as a group by acting like a monopoly, each individual oligopoly faces a private temptation to produce just a slightly higher quantity and earn slightly higher profit—while still counting on the other oligopolists to hold down their production and keep prices high. If at least some oligopolists give in to this temptation and start producing more, then the market price will fall. Indeed, a small handful of oligopoly firms may end up competing so fiercely that they all end up earning zero economic profits—as if they were perfect competitors.

Tradeoffs of Imperfect Competition

Monopolistic competition is probably the single most common market structure in the U.S. economy. It provides powerful incentives for innovation, as firms seek to earn profits in the short run, while entry assures that firms do not earn economic profits in the long run. However,

monopolistically competitive firms do not produce at the lowest point on their average cost curves. In addition, the endless search to impress consumers through product differentiation may lead to excessive social expenses on advertising and marketing.

Oligopoly is probably the second most common market structure. When oligopolies result from patented innovations or from taking advantage of economies of scale to produce at low average cost, they may provide considerable benefit to consumers. Oligopolies are often buffeted by significant barriers to entry, which enable the oligopolists to earn sustained profits over long periods of time. Oligopolists also do not typically produce at the minimum of their average cost curves. When they lack vibrant competition, they may lack incentives to provide innovative products and high-quality service.

The task of public policy with regard to competition is to sort through these multiple realities, attempting to encourage behavior that is beneficial to the broader society and to discourage behavior that only adds to the profits of a few large companies, with no corresponding benefit to consumers. [Monopoly and Antitrust Policy](#) discusses the delicate judgments that go into this task.

Note:**The Temptation to Defy the Law**

Oligopolistic firms have been called “cats in a bag,” as this chapter mentioned. The French detergent makers chose to “cozy up” with each other. The result? An uneasy and tenuous relationship. When the *Wall Street Journal* reported on the matter, it wrote: “According to a statement a Henkel manager made to the [French anti-trust] commission, the detergent makers wanted ‘to limit the intensity of the competition between them and clean up the market.’ Nevertheless, by the early 1990s, a price war had broken out among them.” During the soap executives’ meetings, which sometimes lasted more than four hours, complex pricing structures were established. “One [soap] executive recalled ‘chaotic’ meetings as each side tried to work out how the other had bent the rules.” Like many cartels, the

soap cartel disintegrated due to the very strong temptation for each member to maximize its own individual profits.

How did this soap opera end? After an investigation, French antitrust authorities fined Colgate-Palmolive, Henkel, and Proctor & Gamble a total of €361 million (\$484 million). A similar fate befell the icemakers. Bagged ice is a commodity, a perfect substitute, generally sold in 7- or 22-pound bags. No one cares what label is on the bag. By agreeing to carve up the ice market, control broad geographic swaths of territory, and set prices, the icemakers moved from perfect competition to a monopoly model. After the agreements, each firm was the sole supplier of bagged ice to a region; there were profits in both the long run and the short run. According to the courts: “These companies illegally conspired to manipulate the marketplace.”

Fines totaled about \$600,000—a steep fine considering a bag of ice sells for under \$3 in most parts of the United States.

Even though it is illegal in many parts of the world for firms to set prices and carve up a market, the temptation to earn higher profits makes it extremely tempting to defy the law.

Key Concepts and Summary

An oligopoly is a situation where a few firms sell most or all of the goods in a market. Oligopolists earn their highest profits if they can band together as a cartel and act like a monopolist by reducing output and raising price. Since each member of the oligopoly can benefit individually from expanding output, such collusion often breaks down—especially since explicit collusion is illegal.

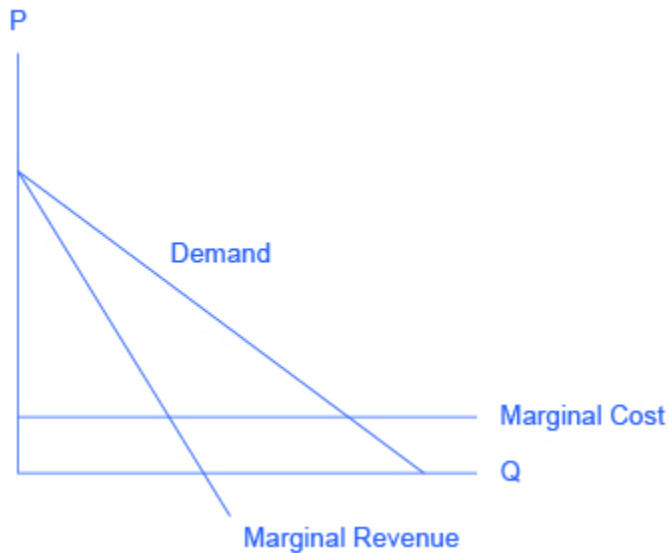
The prisoner’s dilemma is an example of game theory. It shows how, in certain situations, all sides can benefit from cooperative behavior rather than self-interested behavior. However, the challenge for the parties is to find ways to encourage cooperative behavior.

Self-Check Questions

Exercise:

Problem:

Consider the curve shown in [\[link\]](#), which shows the market demand, marginal cost, and marginal revenue curve for firms in an oligopolistic industry. In this example, we assume firms have zero fixed costs.

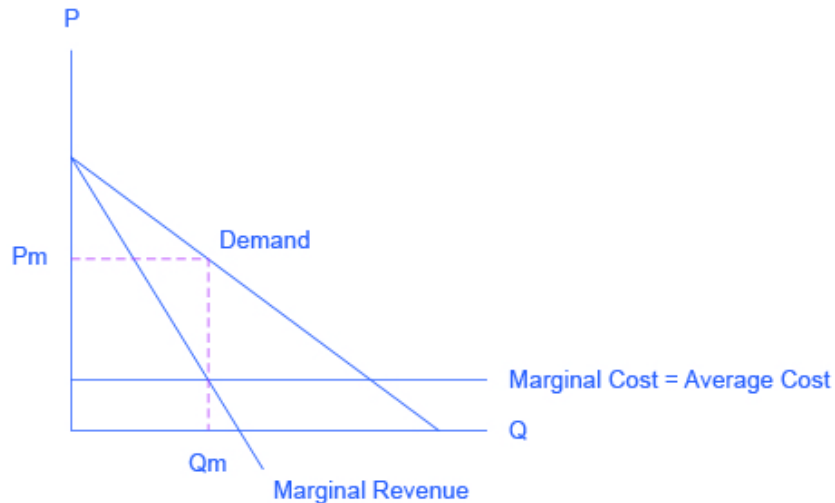


- Suppose the firms collude to form a cartel. What price will the cartel charge? What quantity will the cartel supply? How much profit will the cartel earn?
- Suppose now that the cartel breaks up and the oligopolistic firms compete as vigorously as possible by cutting the price and increasing sales. What will the industry quantity and price be? What will the collective profits be of all firms in the industry?
- Compare the equilibrium price, quantity, and profit for the cartel and cutthroat competition outcomes.

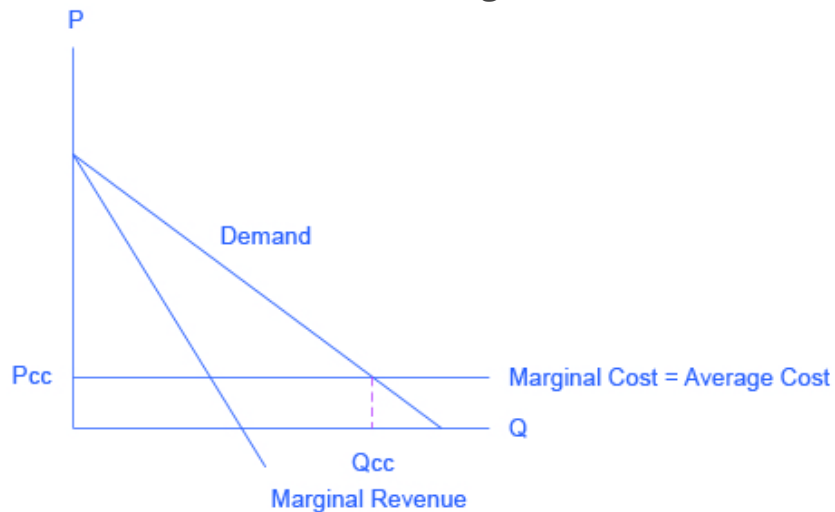
Solution:

- If the firms form a cartel, they will act like a monopoly, choosing the quantity of output where $MR = MC$. Drawing a line from the monopoly quantity up to the demand curve shows the monopoly price. Assuming that fixed costs are zero, and with an understanding of cost and profit, we can infer that when the marginal cost curve is horizontal, average cost is the same as

marginal cost. Thus, the cartel will earn positive economic profits equal to the area of the rectangle, with a base equal to the monopoly quantity and a height equal to the difference between price (on the demand above the monopoly quantity) and average cost, as shown in the following figure.



- b. The firms will expand output and cut price as long as there are profits remaining. The long-run equilibrium will occur at the point where average cost equals demand. As a result, the oligopoly will earn zero economic profits due to “cutthroat competition,” as shown in the next figure.



- c. $P_c > P_{cc}$. $Q_c < Q_{cc}$. Profit for the cartel is positive and large. Profit for cutthroat competition is zero.

Exercise:

Problem:

Sometimes oligopolies in the same industry are very different in size. Suppose we have a duopoly where one firm (Firm A) is large and the other firm (Firm B) is small, as shown in the prisoner's dilemma box in [\[link\]](#).

	Firm B colludes with Firm A	Firm B cheats by selling more output
Firm A colludes with Firm B	A gets \$1,000, B gets \$100	A gets \$800, B gets \$200
Firm A cheats by selling more output	A gets \$1,050, B gets \$50	A gets \$500, B gets \$20

Assuming that the payoffs are known to both firms, what is the likely outcome in this case?

Solution:

Firm B reasons that if it cheats and Firm A does not notice, it will double its money. Since Firm A's profits will decline substantially, however, it is likely that Firm A will notice and if so, Firm A will cheat also, with the result that Firm B will lose 90% of what it gained by cheating. Firm A will reason that Firm B is unlikely to risk cheating. If neither firm cheats, Firm A earns \$1000. If Firm A cheats, assuming Firm B does not cheat, A can boost its profits only a little, since Firm B is so small. If both firms cheat, then Firm A loses at least 50% of what it could have earned. The possibility of a small gain (\$50)

is probably not enough to induce Firm A to cheat, so in this case it is likely that both firms will collude.

Review Questions

Exercise:

Problem:

Will the firms in an oligopoly act more like a monopoly or more like competitors? Briefly explain.

Exercise:

Problem:

Does each individual in a prisoner's dilemma benefit more from cooperation or from pursuing self-interest? Explain briefly.

Exercise:

Problem:

What stops oligopolists from acting together as a monopolist and earning the highest possible level of profits?

Critical Thinking Questions

Exercise:

Problem:

Would you expect the kinked demand curve to be more extreme (like a right angle) or less extreme (like a normal demand curve) if each firm in the cartel produces a near-identical product like OPEC and petroleum? What if each firm produces a somewhat different product? Explain your reasoning.

Exercise:

Problem:

When OPEC raised the price of oil dramatically in the mid-1970s, experts said it was unlikely that the cartel could stay together over the long term—that the incentives for individual members to cheat would become too strong. More than forty years later, OPEC still exists. Why do you think OPEC has been able to beat the odds and continue to collude? *Hint:* You may wish to consider non-economic reasons.

Problems

Exercise:

Problem:

Mary and Raj are the only two growers who provide organically grown corn to a local grocery store. They know that if they cooperated and produced less corn, they could raise the price of the corn. If they work independently, they will each earn \$100. If they decide to work together and both lower their output, they can each earn \$150. If one person lowers output and the other does not, the person who lowers output will earn \$0 and the other person will capture the entire market and will earn \$200. [\[link\]](#) represents the choices available to Mary and Raj. What is the best choice for Raj if he is sure that Mary will cooperate? If Mary thinks Raj will cheat, what should Mary do and why? What is the prisoner’s dilemma result? What is the preferred choice if they could ensure cooperation? A = Work independently; B = Cooperate and Lower Output. (Each results entry lists Raj’s earnings first, and Mary's earnings second.)

	Mary	

		A	B
Raj	A	(\$100, \$100)	(\$200, \$0)
	B	(\$0, \$200)	(\$150, \$150)

Exercise:

Problem:

Jane and Bill are apprehended for a bank robbery. They are taken into separate rooms and questioned by the police about their involvement in the crime. The police tell them each that if they confess and turn the other person in, they will receive a lighter sentence. If they both confess, they will be each be sentenced to 30 years. If neither confesses, they will each receive a 20-year sentence. If only one confesses, the confessor will receive 15 years and the one who stayed silent will receive 35 years. [\[link\]](#) below represents the choices available to Jane and Bill. If Jane trusts Bill to stay silent, what should she do? If Jane thinks that Bill will confess, what should she do? Does Jane have a dominant strategy? Does Bill have a dominant strategy? A = Confess; B = Stay Silent. (Each results entry lists Jane's sentence first (in years), and Bill's sentence second.)

		Jane	
		A	B
Bill	A	(30, 30)	(15, 35)
	B	(35, 15)	(20, 20)

References

The United States Department of Justice. "Antitrust Division." Accessed October 17, 2013. <http://www.justice.gov/atr/>.

eMarketer.com. 2014. "Total US Ad Spending to See Largest Increase Since 2004: Mobile advertising leads growth; will surpass radio, magazines and newspapers this year. Accessed March 12, 2015. <http://www.emarketer.com/Article/Total-US-Ad-Spending-See-Largest-Increase-Since-2004/1010982>.

Federal Trade Commission. "About the Federal Trade Commission." Accessed October 17, 2013. <http://www.ftc.gov/ftc/about.shtm>.

Glossary

cartel

a group of firms that collude to produce the monopoly output and sell at the monopoly price

collusion

when firms act together to reduce output and keep prices high

duopoly

an oligopoly with only two firms

game theory

a branch of mathematics often used by economists that analyzes situations in which players must make decisions and then receive payoffs based on what decisions the other players make

kinked demand curve

a perceived demand curve that arises when competing oligopoly firms commit to match price cuts, but not price increases

prisoner's dilemma

a game in which the gains from cooperation are larger than the rewards from pursuing self-interest

Introduction to Monopoly and Antitrust Policy

class="introduction"

Oligopoly versus Competitors in the Marketplace

Large corporations, such as the natural gas producer Kinder Morgan, can bring economies of scale to the marketplace.

Will that benefit consumers?

Or is more competition better for consumers?

(Credit: modification of work by Derrick Coetzee/Flickr
r Creative Commons)



Note:

More than Cooking, Heating, and Cooling

If you live in the United States, there is a slightly better than 50–50 chance your home is heated and cooled using natural gas. You may even use natural gas for cooking. However, those uses are not the primary uses of natural gas in the U.S. In 2012, according to the U.S. Energy Information Administration, home heating, cooling, and cooking accounted for just 18% of natural gas usage. What accounts for the rest? The greatest uses for natural gas are the generation of electric power (39%) and in industry (30%). Together these three uses for natural gas touch many areas of our lives, so why would there be any opposition to a merger of two natural gas firms? After all, a merger could mean increased efficiencies and reduced costs to people like you and me.

In October 2011, Kinder Morgan and El Paso Corporation, two natural gas firms, announced they were merging. The announcement stated the combined firm would link “nearly every major production region with markets,” cut costs by “eliminating duplication in pipelines and other assets,” and that “the savings could be passed on to consumers.”

The objection? The \$21.1 billion deal would give Kinder Morgan control of more than 80,000 miles of pipeline, making the new firm the third largest energy producer in North America. As the third largest energy producer, policymakers and the public wondered whether the cost savings

really would be passed on to consumers, or would the merger give Kinder Morgan a strong oligopoly position in the natural gas marketplace?

That brings us to the central question this chapter poses: What should the balance be between corporate size and a larger number of competitors in a marketplace? We will also consider what role the government should play in this balancing act.

Note:

Introduction to Monopoly and Antitrust Policy

In this chapter, you will learn about:

- Corporate Mergers
- Regulating Anticompetitive Behavior
- Regulating Natural Monopolies
- The Great Deregulation Experiment

The previous chapters on the theory of the firm identified three important lessons: First, that competition, by providing consumers with lower prices and a variety of innovative products, is a good thing; second, that large-scale production can dramatically lower average costs; and third, that markets in the real world are rarely perfectly competitive. As a consequence, government policymakers must determine how much to intervene to balance the potential benefits of large-scale production against the potential loss of competition that can occur when businesses grow in size, especially through mergers.

For example, in 2011, AT&T and T-Mobile proposed a merger. At the time, there were only four major mobile phone service providers. The proposal was blocked by both the Justice Department and the FCC.

The two companies argued that the merger would benefit consumers, who would be able to purchase better telecommunications services at a cheaper price because the newly created firm would be able to produce more

efficiently by taking advantage of economies of scale and eliminating duplicate investments. However, a number of activist groups like the Consumer Federation of America and Public Knowledge expressed fears that the merger would reduce competition and lead to higher prices for consumers for decades to come. In December 2006, the federal government allowed the merger to proceed. By 2009, the new post-merger AT&T was the eighth largest company by revenues in the United States, and by that measure the largest telecommunications company in the world. Economists have spent – and will still spend – years trying to determine whether the merger of AT&T and BellSouth, as well as other smaller mergers of telecommunications companies at about this same time, helped consumers, hurt them, or did not make much difference.

This chapter discusses public policy issues about competition. How can economists and governments determine when mergers of large companies like AT&T and BellSouth should be allowed and when they should be blocked? The government also plays a role in policing anticompetitive behavior other than mergers, like prohibiting certain kinds of contracts that might restrict competition. In the case of natural monopoly, however, trying to preserve competition probably will not work very well, and so government will often resort to regulation of price and/or quantity of output. In recent decades, there has been a global trend toward less government intervention in the price and output decisions of businesses.

Corporate Mergers

By the end of this section, you will be able to:

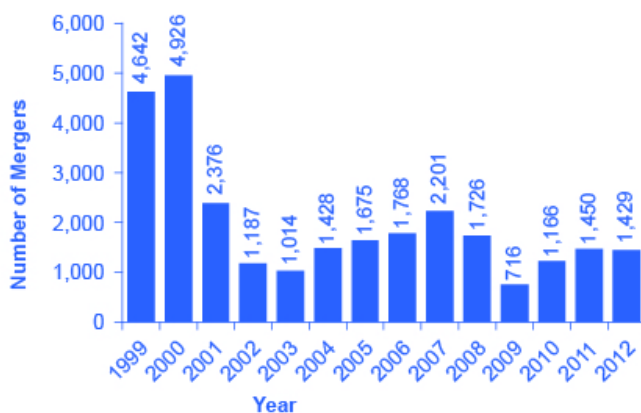
- Explain antitrust law and its significance
- Calculate concentration ratios
- Calculate the Herfindahl-Herschman Index (HHI)
- Evaluate methods of antitrust regulation

A corporate **merger** occurs when two formerly separate firms combine to become a single firm. When one firm purchases another, it is called an **acquisition**. An acquisition may not look just like a merger, since the newly purchased firm may continue to be operated under its former company name. Mergers can also be lateral, where two firms of similar sizes combine to become one. However, both mergers and acquisitions lead to two formerly separate firms being under common ownership, and so they are commonly grouped together.

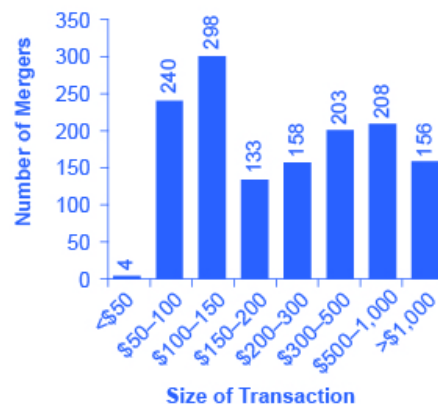
Regulations for Approving Mergers

Since a merger combines two firms into one, it can reduce the extent of competition between firms. Therefore, when two U.S. firms announce a merger or acquisition where at least one of the firms is above a minimum size of sales (a threshold that moves up gradually over time, and was at \$70.9 million in 2013), or certain other conditions are met, they are required under law to notify the U.S. Federal Trade Commission (FTC). The left-hand panel of [\[link\]](#) (a) shows the number of mergers submitted for review to the FTC each year from 1999 to 2012. Mergers were very high in the late 1990s, diminished in the early 2000s, and then rebounded somewhat in a cyclical fashion. The right-hand panel of [\[link\]](#) (b) shows the distribution of those mergers submitted for review in 2012 as measured by the size of the transaction. It is important to remember that this total leaves out many small mergers under \$50 million, which only need to be reported in certain limited circumstances. About a quarter of all reported merger and acquisition transactions in 2012 exceeded \$500 million, while about 11 percent exceeded \$1 billion. In 2014, the FTC took action against mergers likely to stifle competition in markets worth 18.6 billion in sales.

Number and Size of Mergers



(a) Number of mergers submitted for review by the Federal Trade Commission, 1999-2012



(b) Size of transaction for mergers submitted for review in 2012 (in millions of dollars)

(a) The number of mergers in 1999 and 2000 were relatively high compared to the annual numbers seen from 2001–2012. While 2001 and 2007 saw a high number of mergers, these were still only about half the number of mergers in 1999 and 2000. (b) In 2012, the greatest number of mergers submitted for review was for transactions between \$100 and \$150 million.

The laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones, are called **antitrust laws**. Before a large merger happens, the antitrust regulators at the FTC and the U.S. Department of Justice can allow the merger, prohibit it, or allow it if certain conditions are met. One common condition is that the merger will be allowed if the firm agrees to sell off certain parts. For example, in 2006, Johnson & Johnson bought the Pfizer’s “consumer health” division, which included well-known brands like Listerine mouthwash and Sudafed cold medicine. As a condition of allowing the merger, Johnson & Johnson was required to sell off six brands to other firms, including Zantac® heartburn relief medication, Cortizone anti-itch cream, and Balmex diaper rash medication, to preserve a greater degree of competition in these markets.

The U.S. government approves most proposed mergers. In a market-oriented economy, firms have the freedom to make their own choices. Private firms generally have the freedom to:

- expand or reduce production
- set the price they choose
- open new factories or sales facilities or close them
- hire workers or to lay them off
- start selling new products or stop selling existing ones

If the owners want to acquire a firm or be acquired, or to merge with another firm, this decision is just one of many that firms are free to make. In these conditions, the managers of private firms will sometimes make mistakes. They may close down a factory which, it later turns out, would have been profitable. They may start selling a product that ends up losing money. A merger between two companies can sometimes lead to a clash of corporate personalities that makes both firms worse off. But the fundamental belief behind a market-oriented economy is that firms, not governments, are in the best position to know if their actions will lead to attracting more customers or producing more efficiently.

Indeed, government regulators agree that most mergers are beneficial to consumers. As the Federal Trade Commission has noted on its website (as of November, 2013): “Most mergers actually benefit competition and consumers by allowing firms to operate more efficiently.” At the same time, the FTC recognizes, “Some [mergers] are likely to lessen competition. That, in turn, can lead to higher prices, reduced availability of goods or services, lower quality of products, and less innovation. Indeed, some mergers create a concentrated market, while others enable a single firm to raise prices.” The challenge for the antitrust regulators at the FTC and the U.S. Department of Justice is to figure out when a merger may hinder competition. This decision involves both numerical tools and some judgments that are difficult to quantify. The following Clear it Up helps explain how antitrust laws came about.

Note:

What is U.S. antitrust law?

In the closing decades of the 1800s, many industries in the U.S. economy were dominated by a single firm that had most of the sales for the entire country. Supporters of these large firms argued that they could take advantage of economies of scale and careful planning to provide consumers with products at low prices. However, critics pointed out that when competition was reduced, these firms were free to charge more and make permanently higher profits, and that without the goading of competition, it was not clear that they were as efficient or innovative as they could be.

In many cases, these large firms were organized in the legal form of a “trust,” in which a group of formerly independent firms were consolidated together by mergers and purchases, and a group of “trustees” then ran the companies as if they were a single firm. Thus, when the U.S. government passed the **Sherman Antitrust Act** in 1890 to limit the power of these trusts, it was called an antitrust law. In an early demonstration of the law’s power, the U.S. Supreme Court in 1911 upheld the government’s right to break up Standard Oil, which had controlled about 90% of the country’s oil refining, into 34 independent firms, including Exxon, Mobil, Amoco, and Chevron. In 1914, the **Clayton Antitrust Act** outlawed mergers and acquisitions (where the outcome would be to “substantially lessen competition” in an industry), price discrimination (where different customers are charged different prices for the same product), and tied sales (where purchase of one product commits the buyer to purchase some other product). Also in 1914, the Federal Trade Commission (FTC) was created to define more specifically what competition was unfair. In 1950, the **Celler-Kefauver Act** extended the Clayton Act by restricting vertical and conglomerate mergers. In the twenty-first century, the FTC and the U.S. Department of Justice continue to enforce antitrust laws.

The Four-Firm Concentration Ratio

Regulators have struggled for decades to measure the degree of monopoly power in an industry. An early tool was the **concentration ratio**, which measures what share of the total sales in the industry are accounted for by

the largest firms, typically the top four to eight firms. For an explanation of how high market concentrations can create inefficiencies in an economy, refer to [Monopoly](#).

Say that the market for replacing broken automobile windshields in a certain city has 18 firms with the market shares shown in [\[link\]](#), where the **market share** is each firm’s proportion of total sales in that market. The four-firm concentration ratio is calculated by adding the market shares of the four largest firms: in this case, $16 + 10 + 8 + 6 = 40$. This concentration ratio would not be considered especially high, because the largest four firms have less than half the market.

If the market shares in the market for replacing automobile windshields are:	
Smooth as Glass Repair Company	16% of the market
The Auto Glass Doctor Company	10% of the market
Your Car Shield Company	8% of the market
Seven firms that each have 6% of the market	42% of the market, combined
Eight firms that each have 3% of the market	24% of the market, combined
Then the four-firm concentration ratio is $16 + 10 + 8 + 6 = 40$.	

Calculating Concentration Ratios from Market Shares

The concentration ratio approach can help to clarify some of the fuzziness over deciding when a merger might affect competition. For instance, if two

of the smallest firms in the hypothetical market for repairing automobile windshields merged, the four-firm concentration ratio would not change—which implies that there is not much worry that the degree of competition in the market has notably diminished. However, if the top two firms merged, then the four-firm concentration ratio would become 46 (that is, $26 + 8 + 6$). While this concentration ratio is modestly higher, the four-firm concentration ratio would still be less than half, so such a proposed merger might barely raise an eyebrow among antitrust regulators.

Note:

Visit this [website](#) to read an article about Google’s run-in with the FTC.



The Herfindahl-Hirshman Index

A four-firm concentration ratio is a simple tool, which may reveal only part of the story. For example, consider two industries that both have a four-firm concentration ratio of 80. However, in one industry five firms each control 20% of the market, while in the other industry, the top firm holds 77% of the market and all the other firms have 1% each. Although the four-firm concentration ratios are identical, it would be reasonable to worry more about the extent of competition in the second case—where the largest firm is nearly a monopoly—than in the first.

Another approach to measuring industry concentration that can distinguish between these two cases is called the **Herfindahl-Hirschman Index (HHI)**. The HHI, as it is often called, is calculated by summing the squares

of the market share of each firm in the industry, as the following Work it Out shows.

Note:

Calculating HHI

Step 1. Calculate the HHI for a monopoly with a market share of 100%. Because there is only one firm, it has 100% market share. The HHI is $100^2 = 10,000$.

Step 2. For an extremely competitive industry, with dozens or hundreds of extremely small competitors, the value of the HHI might drop as low as 100 or even less. Calculate the HHI for an industry with 100 firms that each have 1% of the market. In this case, the HHI is $100(1^2) = 100$.

Step 3. Calculate the HHI for the industry shown in [\[link\]](#). In this case, the HHI is $16^2 + 10^2 + 8^2 + 7(6^2) + 8(3^2) = 744$.

Step 4. Note that the HHI gives greater weight to large firms.

Step 5. Consider the example given earlier, comparing one industry where five firms each have 20% of the market with an industry where one firm has 77% and the other 23 firms have 1% each. The two industries have the same four-firm concentration ratio of 80. But the HHI for the first industry is $5(20^2) = 2,000$, while the HHI for the second industry is much higher at $77^2 + 23(1^2) = 5,952$.

Step 6. Note that the near-monopolist in the second industry drives up the HHI measure of industrial concentration.

Step 7. Review [\[link\]](#) which gives some examples of the four-firm concentration ratio and the HHI in various U.S. industries in 2009. (You can find market share data from multiple industry sources. Data in the table are from: Verizon (for wireless), *The Wall Street Journal* (for automobiles), IDC Worldwide (for computers) and the U.S. Bureau of Transportation Statistics (for airlines).)

U.S. Industry	Four-Firm Ratio	HHI
<i>Wireless</i>	91	2,311
Largest five: Verizon, AT&T, Sprint, T-Mobile, MetroPCS		
<i>Automobiles</i>	63	1,121
Largest five: GM, Toyota, Ford, Honda, Chrysler		
<i>Computers</i>	74	1,737
Largest five: HP, Dell, Acer, Apple, Toshiba		
<i>Airlines</i>	44	536
Largest five: Southwest, American, Delta, United, U.S. Airways		
Examples of Concentration Ratios and HHIs in the U.S. Economy, 2009		

In the 1980s, the FTC followed these guidelines: If a merger would result in an HHI of less than 1,000, the FTC would probably approve it. If a merger would result in an HHI of more than 1,800, the FTC would probably challenge it. If a merger would result in an HHI between 1,000 and 1,800, then the FTC would scrutinize the plan and make a case-by-case decision. However, in the last several decades, the antitrust enforcement authorities have moved away from relying as heavily on measures of concentration ratios and HHIs to determine whether a merger will be allowed, and instead

carried out more case-by-case analysis on the extent of competition in different industries.

New Directions for Antitrust

Both the four-firm concentration ratio and the Herfindahl-Hirschman index share some weaknesses. First, they begin from the assumption that the “market” under discussion is well-defined, and the only question is measuring how sales are divided in that market. Second, they are based on an implicit assumption that competitive conditions across industries are similar enough that a broad measure of concentration in the market is enough to make a decision about the effects of a merger. These assumptions, however, are not always correct. In response to these two problems, the antitrust regulators have been changing their approach in the last decade or two.

Defining a **market** is often controversial. For example, Microsoft in the early 2000s had a dominant share of the software for computer operating systems. However, in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014. A narrowly defined market will tend to make concentration appear higher, while a broadly defined market will tend to make it appear smaller.

There are two especially important shifts affecting how markets are defined in recent decades: one centers on technology and the other centers on globalization. In addition, these two shifts are interconnected. With the vast improvement in communications technologies, including the development of the Internet, a consumer can order books or pet supplies from all over the country or the world. As a result, the degree of competition many local retail businesses face has increased. The same effect may operate even more strongly in markets for business supplies, where so-called “business-to-business” websites can allow buyers and suppliers from anywhere in the world to find each other.

Globalization has changed the boundaries of markets. As recently as the 1970s, it was common for measurements of concentration ratios and HHIs

to stop at national borders. Now, many industries find that their competition comes from the global market. A few decades ago, three companies, General Motors, Ford, and Chrysler, dominated the U.S. auto market. By 2014, however, these three firms were making less than half of U.S. auto sales, and facing competition from well-known car manufacturers such as Toyota, Honda, Nissan, Volkswagen, Mitsubishi, and Mazda. When HHIs are calculated with a global perspective, concentration in most major industries—including cars—is lower than in a purely domestic context.

Because attempting to define a particular market can be difficult and controversial, the Federal Trade Commission has begun to look less at market share and more at the data on actual competition between businesses. For example, in February 2007, Whole Foods Market and Wild Oats Market announced that they wished to merge. These were the two largest companies in the market that the government defined as “premium natural and organic supermarket chains.” However, one could also argue that they were two relatively small companies in the broader market for all stores that sell groceries or specialty food products.

Rather than relying on a market definition, the government antitrust regulators looked at detailed evidence on profits and prices for specific stores in different cities, both before and after other competitive stores entered or exited. Based on that evidence, the Federal Trade Commission decided to block the merger. After two years of legal battles, the merger was eventually allowed in 2009 under the conditions that Whole Foods sell off the Wild Oats brand name and a number of individual stores, to preserve competition in certain local markets. For more on the difficulties of defining markets, refer to [Monopoly](#).

This new approach to antitrust regulation involves detailed analysis of specific markets and companies, instead of defining a market and counting up total sales. A common starting point is for antitrust regulators to use statistical tools and real-world evidence to estimate the **demand curves** and **supply curves** faced by the firms that are proposing the merger. A second step is to specify how competition occurs in this specific industry. Some possibilities include competing to cut prices, to raise output, to build a brand name through advertising, and to build a reputation for good service

or high quality. With these pieces of the puzzle in place, it is then possible to build a statistical model that estimates the likely outcome for consumers if the two firms are allowed to merge. Of course, these models do require some degree of subjective judgment, and so they can become the subject of legal disputes between the antitrust authorities and the companies that wish to merge.

Key Concepts and Summary

A corporate merger involves two private firms joining together. An acquisition refers to one firm buying another firm. In either case, two formerly independent firms become one firm. Antitrust laws seek to ensure active competition in markets, sometimes by preventing large firms from forming through mergers and acquisitions, sometimes by regulating business practices that might restrict competition, and sometimes by breaking up large firms into smaller competitors.

A four-firm concentration ratio is one way of measuring the extent of competition in a market. It is calculated by adding the market shares—that is, the percentage of total sales—of the four largest firms in the market. A Herfindahl-Hirschman Index (HHI) is another way of measuring the extent of competition in a market. It is calculated by taking the market shares of all firms in the market, squaring them, and then summing the total.

The forces of globalization and new communications and information technology have increased the level of competition faced by many firms by increasing the amount of competition from other regions and countries.

Self-Check Questions

Exercise:

Problem:

Is it true that both the four-firm concentration ratio and the Herfindahl-Hirschman Index can be affected by a merger between two firms that are not already in the top four by size? Explain briefly.

Solution:

Yes, it is true. The HHI example is easy enough: since the market shares of all firms are included in the HHI calculation, a merger between two of the firms will change the HHI. For the four-firm concentration ratio, it is quite possible that a merger between, say, the fifth and sixth largest firms in the market could create a new firm that is then ranked in the top four in the market. In this case, a merger of two firms, neither in the top four, would still change the four-firm concentration ratio.

Exercise:**Problem:**

Is it true that the four-firm concentration ratio puts more emphasis on one or two very large firms, while the Herfindahl-Hirshman Index puts more emphasis on all the firms in the entire market? Explain briefly.

Solution:

No, it is not true. The HHI includes the market shares of all firms in its calculation, but the squaring of the market shares has the effect of making the impact of the largest firms relatively bigger than in the 4-firm or 8-firm ratio.

Exercise:**Problem:**

Some years ago, two intercity bus companies, Greyhound Lines, Inc. and Trailways Transportation System, wanted to merge. One possible definition of the market in this case was “the market for intercity bus service.” Another possible definition was “the market for intercity transportation, including personal cars, car rentals, passenger trains, and commuter air flights.” Which definition do you think the bus companies preferred, and why?

Solution:

The bus companies wanted the broader market definition (i.e., the second definition). If the narrow definition had been used, the combined bus companies would have had a near-monopoly on the market for intercity bus service. But they had only a sliver of the market for intercity transportation when everything else was included. The merger was allowed.

Exercise:

Problem:

As a result of globalization and new information and communications technology, would you expect that the definitions of markets used by antitrust authorities will become broader or narrower?

Solution:

The common expectation is that the definition of markets will become broader because of greater competition from faraway places. However, this broadening doesn't necessarily mean that antitrust authorities can relax. There is also a fear that companies with a local or national monopoly may use the new opportunities to extend their reach across national borders, and that it will be difficult for national authorities to respond.

Review Questions

Exercise:

Problem: What is a corporate merger? What is an acquisition?

Exercise:

Problem: What is the goal of antitrust policies?

Exercise:

Problem:

How is a four-firm concentration ratio measured? What does a high measure mean about the extent of competition?

Exercise:**Problem:**

How is a Herfindahl-Hirshman Index measured? What does a low measure mean about the extent of competition?

Exercise:**Problem:**

Why can it be difficult to decide what a “market” is for purposes of measuring competition?

Critical Thinking Questions**Exercise:****Problem:**

Does either the four-firm concentration ratio or the HHI directly measure the amount of competition in an industry? Why or why not?

Exercise:**Problem:**

What would be evidence of serious competition between firms in an industry? Can you identify two highly competitive industries?

Problems**Exercise:**

Problem:

Use [\[link\]](#) to calculate the four-firm concentration ratio for the U.S. auto market. Does this indicate a concentrated market or not?

GM	19%
Ford	17%
Toyota	14%
Chrysler	11%

Global Auto Manufacturers with Top Four U.S. Market Share, June 2013(Source: <http://www.zacks.com/commentary/27690/auto-industry-stock-outlook-june-2013>)

Exercise:**Problem:**

Use [\[link\]](#) and [\[link\]](#) to calculate the Herfindal-Hirschman Index for the U.S. auto market. Would the FTC approve a merger between GM and Ford?

Honda	10%
Nissan	7%

Hyundai	5%
Kia	4%
Subaru	3%
Volkswagen	3%

Global Auto Manufacturers with additional U.S. Market Share, June 2013(Source: <http://www.zacks.com/commentary/27690/auto-industry-stock-outlook-june-2013>)

Glossary

acquisition

when one firm purchases another

antitrust laws

laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones

concentration ratio

an early tool to measure the degree of monopoly power in an industry; measures what share of the total sales in the industry are accounted for by the largest firms, typically the top four to eight firms

four-firm concentration ratio

the percentage of the total sales in the industry that are accounted for by the largest four firms

Herfindahl-Hirschman Index (HHI)

approach to measuring market concentration by adding the square of the market share of each firm in the industry

market share

the percentage of total sales in the market

merger

when two formerly separate firms combine to become a single firm

Regulating Anticompetitive Behavior

By the end of this section, you will be able to:

- Analyze restrictive practices
- Explain tying sales, bundling, and predatory pricing
- Evaluate a real-world situation of possible anticompetitive and restrictive practices

The U.S. antitrust laws reach beyond blocking mergers that would reduce competition to include a wide array of anticompetitive practices. For example, it is illegal for competitors to form a cartel to collude to make pricing and output decisions, as if they were a monopoly firm. The Federal Trade Commission and the U.S. Department of Justice prohibit firms from agreeing to fix prices or output, rigging bids, or sharing or dividing markets by allocating customers, suppliers, territories, or lines of commerce.

In the late 1990s, for example, the antitrust regulators prosecuted an international cartel of vitamin manufacturers, including the Swiss firm Hoffman-La Roche, the German firm BASF, and the French firm Rhone-Poulenc. These firms reached agreements on how much to produce, how much to charge, and which firm would sell to which customers. The high-priced vitamins were then bought by firms like General Mills, Kellogg, Purina-Mills, and Proctor and Gamble, which pushed up the prices more. Hoffman-La Roche pleaded guilty in May 1999 and agreed both to pay a fine of \$500 million and to have at least one top executive serve four months of jail time.

Under U.S. antitrust laws, monopoly itself is not illegal. If a firm has a monopoly because of a newly patented invention, for example, the law explicitly allows a firm to earn higher-than-normal profits for a time as a reward for innovation. If a firm achieves a large share of the market by producing a better product at a lower price, such behavior is not prohibited by antitrust law.

Restrictive Practices

Antitrust law includes rules against **restrictive practices**—practices that do not involve outright agreements to raise price or to reduce the quantity produced, but that might have the effect of reducing competition. Antitrust cases involving restrictive practices are often controversial, because they delve into specific contracts or agreements between firms that are allowed in some cases but not in others.

For example, if a product manufacturer is selling to a group of dealers who then sell to the general public it is illegal for the manufacturer to demand a **minimum resale price maintenance agreement**, which would require the dealers to sell for at least a certain minimum price. A minimum price contract is illegal because it would restrict competition among dealers. However, the manufacturer is legally allowed to “suggest” minimum prices and to stop selling to dealers who regularly undercut the suggested price. If you think this rule sounds like a fairly subtle distinction, you are right.

An **exclusive dealing** agreement between a manufacturer and a dealer can be legal or illegal. It is legal if the purpose of the contract is to encourage competition between dealers. For example, it is legal for the Ford Motor Company to sell its cars to only Ford dealers, for General Motors to sell to only GM dealers, and so on. However, exclusive deals may also limit competition. If one large retailer obtained the exclusive rights to be the sole distributor of televisions, computers, and audio equipment made by a number of companies, then this exclusive contract would have an anticompetitive effect on other retailers.

Tying sales happen when a customer is required to buy one product only if the customer also buys a second product. Tying sales are controversial because they force consumers to purchase a product that they may not actually want or need. Further, the additional, required products are not necessarily advantageous to the customer. Suppose that to purchase a popular DVD, the store required that you also purchase a portable TV of a certain model. These products are only loosely related, thus there is no reason to make the purchase of one contingent on the other. Even if a customer was interested in a portable TV, the tying to a particular model prevents the customer from having the option of selecting one from the numerous types available in the market. A related, but not identical, concept

is called **bundling**, where two or more products are sold as one. Bundling typically offers an advantage for the consumer by allowing them to acquire multiple products or services for a better price. For example, several cable companies allow customers to buy products like cable, internet, and a phone line through a special price available through bundling. Customers are also welcome to purchase these products separately, but the price of bundling is usually more appealing.

In some cases, tying sales and bundling can be viewed as anticompetitive. However, in other cases they may be legal and even common. It is common for people to purchase season tickets to a sports team or a set of concerts so that they can be guaranteed tickets to the few contests or shows that are most popular and likely to sell out. Computer software manufacturers may often bundle together a number of different programs, even when the buyer wants only a few of the programs. Think about the software that is included in a new computer purchase, for example.

Recall from the chapter on [Monopoly](#) that predatory pricing occurs when the existing firm (or firms) reacts to a new firm by dropping prices very low, until the new firm is driven out of the market, at which point the existing firm raises prices again. This pattern of pricing is aimed at deterring the entry of new firms into the market. But in practice, it can be hard to figure out when pricing should be considered predatory. Say that American Airlines is flying between two cities, and a new airline starts flying between the same two cities, at a lower price. If American Airlines cuts its price to match the new entrant, is this predatory pricing? Or is it just market competition at work? A commonly proposed rule is that if a firm is selling for less than its average variable cost—that is, at a price where it should be shutting down—then there is evidence for predatory pricing. But calculating in the real world what costs are variable and what costs are fixed is often not obvious, either.

The Microsoft antitrust case embodies many of these gray areas in restrictive practices, as the next Clear it Up shows.

Note:

Did Microsoft® engage in anticompetitive and restrictive practices?

The most famous restrictive practices case of recent years was a series of lawsuits by the U.S. government against Microsoft—lawsuits that were encouraged by some of Microsoft’s competitors. All sides admitted that Microsoft’s Windows program had a near-monopoly position in the market for the software used in general computer operating systems. All sides agreed that the software had many satisfied customers. All sides agreed that the capabilities of computer software that was compatible with Windows—both software produced by Microsoft and that produced by other companies—had expanded dramatically in the 1990s. Having a **monopoly** or a near-monopoly is not necessarily illegal in and of itself, but in cases where one company controls a great deal of the market, antitrust regulators look at any allegations of restrictive practices with special care. The antitrust regulators argued that Microsoft had gone beyond profiting from its software innovations and its dominant position in the software market for operating systems, and had tried to use its market power in operating systems software to take over other parts of the software industry. For example, the government argued that Microsoft had engaged in an anticompetitive form of exclusive dealing by threatening computer makers that, if they did not leave another firm’s software off their machines (specifically, Netscape’s Internet browser), then Microsoft would not sell them its operating system software. Microsoft was accused by the government antitrust regulators of tying together its Windows operating system software, where it had a monopoly, with its Internet Explorer browser software, where it did not have a monopoly, and thus using this bundling as an anticompetitive tool. Microsoft was also accused of a form of predatory pricing; namely, giving away certain additional software products for free as part of Windows, as a way of driving out the competition from other makers of software.

In April 2000, a federal court held that Microsoft’s behavior had crossed the line into unfair competition, and recommended that the company be broken into two competing firms. However, that penalty was overturned on appeal, and in November 2002 Microsoft reached a settlement with the government that it would end its restrictive practices.

The concept of restrictive practices is continually evolving, as firms seek new ways to earn profits and government regulators define what is permissible and what is not. A situation where the law is evolving and changing is always somewhat troublesome, since laws are most useful and fair when firms know what they are in advance. In addition, since the law is open to interpretation, competitors who are losing out in the market can accuse successful firms of anticompetitive restrictive practices, and try to win through government regulation what they have failed to accomplish in the market. Officials at the Federal Trade Commission and the Department of Justice are, of course, aware of these issues, but there is no easy way to resolve them.

Key Concepts and Summary

Firms are blocked by antitrust authorities from openly colluding to form a cartel that will reduce output and raise prices. Companies sometimes attempt to find other ways around these restrictions and, consequently, many antitrust cases involve restrictive practices that can reduce competition in certain circumstances, like tie-in sales, bundling, and predatory pricing.

Self-Check Question

Exercise:

Problem:

Why would a firm choose to use one or more of the anticompetitive practices described in [Regulating Anticompetitive Behavior](#)?

Solution:

Because outright collusion to raise profits is illegal and because existing regulations include gray areas which firms may be able to exploit.

Review Questions

Exercise:

Problem:

What is a minimum resale price maintenance agreement? How might it reduce competition and when might it be acceptable?

Exercise:

Problem:

What is exclusive dealing? How might it reduce competition and when might it be acceptable?

Exercise:

Problem:

What is a tie-in sale? How might it reduce competition and when might it be acceptable?

Exercise:

Problem:

What is predatory pricing? How might it reduce competition, and why might it be difficult to tell when it should be illegal?

Critical Thinking Questions

Exercise:

Problem:

Can you think of any examples of successful predatory pricing in the real world?

Exercise:

Problem:

If you were developing a product (like a web browser) for a market with significant barriers to entry, how would you try to get your product into the market successfully?

Glossary**bundling**

a situation in which multiple products are sold as one

exclusive dealing

an agreement that a dealer will sell only products from one manufacturer

minimum resale price maintenance agreement

an agreement that requires a dealer who buys from a manufacturer to sell for at least a certain minimum price

restrictive practices

practices that reduce competition but that do not involve outright agreements between firms to raise prices or to reduce the quantity produced

tying sales

a situation where a customer is allowed to buy one product only if the customer also buys another product

Regulating Natural Monopolies

By the end of this section, you will be able to:

- Evaluate the appropriate competition policy for a natural monopoly
- Interpret a graph of regulatory choices
- Contrast cost-plus and price cap regulation

Most true monopolies today in the U.S. are regulated, natural monopolies. A natural monopoly poses a difficult challenge for competition policy, because the structure of costs and demand seems to make competition unlikely or costly. A **natural monopoly** arises when average costs are declining over the range of production that satisfies market demand. This typically happens when fixed costs are large relative to variable costs. As a result, one firm is able to supply the total quantity demanded in the market at lower cost than two or more firms—so splitting up the natural monopoly would raise the average cost of production and force customers to pay more.

Public utilities, the companies that have traditionally provided water and electrical service across much of the United States, are leading examples of natural monopoly. It would make little sense to argue that a local water company should be broken up into several competing companies, each with its own separate set of pipes and water supplies. Installing four or five identical sets of pipes under a city, one for each water company, so that each household could choose its own water provider, would be terribly costly. The same argument applies to the idea of having many competing companies for delivering electricity to homes, each with its own set of wires. Before the advent of wireless phones, the argument also applied to the idea of many different phone companies, each with its own set of phone wires running through the neighborhood.

Cost-Plus versus Price Cap Regulation

Indeed, regulators of public utilities for many decades followed the general approach of attempting to choose a price. They calculated the average cost of production for the water or electricity companies, added in an amount for the normal rate of profit the firm should expect to earn, and set the price for consumers accordingly. This method was known as **cost-plus regulation**.

Cost-plus regulation raises difficulties of its own. If producers are reimbursed for their costs, plus a bit more, then at a minimum, producers have less reason to be concerned with high costs—because they can just pass them along in higher prices. Worse, firms under cost-plus regulation even have an incentive to generate high costs by building huge factories or employing lots of staff, because what they can charge is linked to the costs they incur.

Thus, in the 1980s and 1990s, some regulators of public utilities began to use **price cap regulation**, where the regulator sets a price that the firm can charge over the next few years. A common pattern was to require a price that declined slightly over time. If the firm can find ways of reducing its costs more quickly than the price caps, it can make a high level of profits. However, if the firm cannot keep up with the price caps or suffers bad luck in the market, it may suffer losses. A few years down the road, the regulators will then set a new series of price caps based on the firm's performance.

Price cap regulation requires delicacy. It will not work if the price regulators set the price cap unrealistically low. It may not work if the market changes dramatically so that the firm is doomed to incurring losses no matter what it does—say, if energy prices rise dramatically on world markets, then the company selling natural gas or heating oil to homes may not be able to meet price caps that seemed reasonable a year or two ago. But if the regulators compare the prices with producers of the same good in other areas, they can, in effect, pressure a natural monopoly in one area to compete with the prices being charged in other areas. Moreover, the possibility of earning greater profits or experiencing losses—instead of having an average rate of profit locked in every year by cost-plus regulation—can provide the natural monopoly with incentives for efficiency and innovation.

With natural monopoly, market competition is unlikely to take root, so if consumers are not to suffer the high prices and restricted output of an unrestricted monopoly, government regulation will need to play a role. In attempting to design a system of price cap regulation with flexibility and incentive, government regulators do not have an easy task.

Key Concepts and Summary

In the case of a natural monopoly, market competition will not work well and so, rather than allowing an unregulated monopoly to raise price and reduce output, the government may wish to regulate price and/or output. Common examples of regulation are public utilities, the regulated firms that often provide electricity and water service.

Cost-plus regulation refers to government regulation of a firm which sets the price that a firm can charge over a period of time by looking at the firm's accounting costs and then adding a normal rate of profit. Price cap regulation refers to government regulation of a firm where the government sets a price level several years in advance. In this case, the firm can either make high profits if it manages to produce at lower costs or sell a higher quantity than expected or suffer low profits or losses if costs are high or it sells less than expected.

Self-Check Questions

Exercise:

Problem:

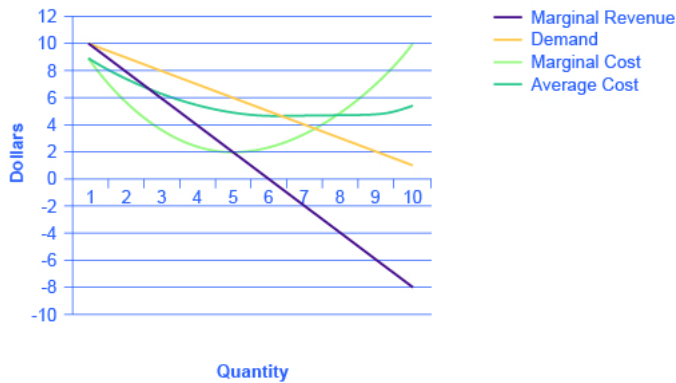
Urban transit systems, especially those with rail systems, typically experience significant economies of scale in operation. Consider the transit system whose data is given in the [link](#). Note that the quantity is in millions of riders.

Demand:	Quantity	1	2	3	4	5	6	7	8	9	10
	Price	10	9	8	7	6	5	4	3	2	1
	Marginal Revenue	10	8	6	4	2	0	-2	-4	-6	-8
Costs:	Marginal Cost	9	6	5	3	2	3	4	5	7	10
	Average Cost	9	7.5	6.7	5.8	5	4.7	4.6	4.6	4.9	5.4

Draw the demand, marginal revenue, marginal cost, and average cost curves. Do they have the normal shapes?

Solution:

Yes, all curves have normal shapes.



Exercise:

Problem:

From the graph you drew to answer [\[link\]](#), would you say this transit system is a natural monopoly? Justify.

Solution:

Yes it is a natural monopoly because average costs decline over the range that satisfies the market demand. For example, at the point where the demand curve and the average cost curve meet, there are economies of scale.

Review Questions

Exercise:

Problem: If public utilities are a natural monopoly, what would be the danger in deregulating them?

Exercise:

Problem:

If public utilities are a natural monopoly, what would be the danger in splitting them up into a number of separate competing firms?

Exercise:

Problem: What is cost-plus regulation?

Exercise:

Problem: What is price cap regulation?

Critical Thinking Questions

Exercise:

Problem:

In the middle of the twentieth century, major U.S. cities had multiple competing city bus companies. Today, there is usually only one and it runs as a subsidized, regulated monopoly. What do you suppose caused the change?

Exercise:

Problem:

Why are urban areas willing to subsidize urban transit systems? Does the argument for subsidies make sense to you?

Problems

Use [\[link\]](#) to answer the following questions.

Exercise:**Problem:**

If the transit system was allowed to operate as an unregulated monopoly, what output would it supply and what price would it charge?

Exercise:**Problem:**

If the transit system was regulated to operate with no subsidy (i.e., at zero economic profit), what approximate output would it supply and what approximate price would it charge?

Exercise:**Problem:**

If the transit system was regulated to provide the most allocatively efficient quantity of output, what output would it supply and what price would it charge? What subsidy would be necessary to insure this efficient provision of transit services?

Glossary

cost-plus regulation

when regulators permit a regulated firm to cover its costs and to make a normal level of profit

price cap regulation

when the regulator sets a price that a firm cannot exceed over the next few years

The Great Deregulation Experiment

By the end of this section, you will be able to:

- Evaluate the effectiveness of price regulation and antitrust policy
- Explain regulatory capture and its significance

Governments at all levels across the United States have regulated prices in a wide range of industries. In some cases, like water and electricity that have natural monopoly characteristics, there is some room in economic theory for such regulation. But once politicians are given a basis to intervene in markets and to choose prices and quantities, it is hard to know where to stop.

Doubts about Regulation of Prices and Quantities

Beginning in the 1970s, it became clear to policymakers of all political leanings that the existing price regulation was not working well. The United States carried out a great policy experiment—the **deregulation** discussed in [Monopoly](#)—removing government controls over prices and quantities produced in airlines, railroads, trucking, intercity bus travel, natural gas, and bank interest rates. The Clear it Up discusses the outcome of deregulation in one industry in particular—airlines.

Note:

What are the results of airline deregulation?

Why did the pendulum swing in favor of deregulation? Consider the airline industry. In the early days of air travel, no airline could make a profit just by flying passengers. Airlines needed something else to carry and the Postal Service provided that something with airmail. And so the first U.S. government regulation of the airline industry happened through the Postal Service, when in 1926 the Postmaster General began giving airlines permission to fly certain routes based on the needs of mail delivery—and the airlines took some passengers along for the ride. In 1934, the Postmaster General was charged by the antitrust authorities with colluding

with the major airlines of that day to monopolize the nation's airways. In 1938, the Civil Aeronautics Board (CAB) was created to regulate airfares and routes instead. For 40 years, from 1938 to 1978, the CAB approved all fares, controlled all entry and exit, and specified which airlines could fly which routes. There was zero entry of new airlines on the main routes across the country for 40 years, because the CAB did not think it was necessary.

In 1978, the Airline Deregulation Act took the government out of the business of determining airfares and schedules. The new law shook up the industry. Famous old airlines like Pan American, Eastern, and Braniff went bankrupt and disappeared. Some new airlines like People Express were created—and then vanished.

The greater competition from deregulation reduced airfares by about one-third over the next two decades, saving consumers billions of dollars a year. The average flight used to take off with just half its seats full; now it is two-thirds full, which is far more efficient. Airlines have also developed hub-and-spoke systems, where planes all fly into a central hub city at a certain time and then depart. As a result, one can fly between any of the spoke cities with just one connection—and there is greater service to more cities than before deregulation. With lower fares and more service, the number of air passengers doubled from the late 1970s to the start of the 2000s—an increase that, in turn, doubled the number of jobs in the airline industry. Meanwhile, with the watchful oversight of government safety inspectors, commercial air travel has continued to get safer over time. The U.S. airline industry is far from perfect. For example, a string of mergers in recent years has raised concerns over how competition might be compromised.

One difficulty with government price regulation is what economists call **regulatory capture**, in which the firms supposedly being regulated end up playing a large role in setting the regulations that they will follow. When the airline industry was being regulated, for example, it suggested appointees to the regulatory board, sent lobbyists to argue with the board, provided most of the information on which the board made decisions, and offered well-paid jobs to at least some of the people leaving the board. In this situation,

consumers can easily end up being not very well represented by the regulators. The result of regulatory capture is that government price regulation can often become a way for existing competitors to work together to reduce output, keep prices high, and limit competition.

The Effects of Deregulation

Deregulation, both of airlines and of other industries, has its negatives. The greater pressure of competition led to entry and exit. When firms went bankrupt or contracted substantially in size, they laid off workers who had to find other jobs. Market competition is, after all, a full-contact sport.

A number of major accounting scandals involving prominent corporations such as Enron, Tyco International, and WorldCom led to the **Sarbanes-Oxley Act** in 2002. Sarbanes-Oxley was designed to increase confidence in financial information provided by public corporations to protect investors from accounting fraud.

The Great Recession which began in late 2007 and which the U.S. economy is still struggling to recover from was caused at least in part by a global financial crisis, which began in the United States. The key component of the crisis was the creation and subsequent failure of several types of unregulated financial assets, such as collateralized mortgage obligations (CMOs, a type of mortgage-backed security), and credit default swaps (CDSs, insurance contracts on assets like CMOs that provided a payoff even if the holder of the CDS did not own the CMO). Many of these assets were rated very safe by private credit rating agencies such as Standard & Poors, Moody's, and Fitch.

The collapse of the markets for these assets precipitated the financial crisis and led to the failure of Lehman Brothers, a major investment bank, numerous large commercial banks, such as Wachovia, and even the Federal National Mortgage Corporation (Fannie Mae), which had to be nationalized—that is, taken over by the federal government. One response to the financial crisis was the **Dodd-Frank Act**, which attempted major reforms of the financial system. The legislation's purpose, as noted on dodd-frank.com is:

"To promote the financial stability of the United States by improving accountability and transparency in the financial system, to end "too big to fail," to protect the American taxpayer by ending bailouts, [and] to protect consumers from abusive financial services practices. . . "

We will explore the financial crisis and the Great Recession in more detail in the macroeconomic chapters of this book, but for now it should be clear that many Americans have grown disenchanted with deregulation, at least of financial markets.

All market-based economies operate against a background of laws and regulations, including laws about enforcing contracts, collecting taxes, and protecting health and the environment. The government policies discussed in this chapter—like blocking certain anticompetitive mergers, ending restrictive practices, imposing price cap regulation on natural monopolies, and deregulation—demonstrate the role of government to strengthen the incentives that come with a greater degree of competition.

Note:

More than Cooking, Heating, and Cooling

What did the Federal Trade Commission (FTC) decide on the Kinder Morgan / El Paso Corporation merger? After careful examination, federal officials decided there was only one area of significant overlap that might provide the merged firm with strong market power. The FTC approved the merger, provided Kinder Morgan divest itself of the overlap area. Tallgrass purchased Kinder Morgan Interstate Gas Transmission, Trailblazer Pipeline Co. LLC, two processing facilities in Wyoming, and Kinder Morgan's 50 percent interest in the Rockies Express Pipeline to meet the FTC requirements. The FTC was attempting to strike a balance between potential cost reductions resulting from economies of scale and concentration of market power.

Did the price of natural gas decrease? Yes, rather significantly. In 2010, the wellhead price of natural gas was \$4.48 per thousand cubic foot; in 2012 the price had fallen to just \$2.66. Was the merger responsible for the large drop in price? The answer is uncertain. The larger contributor to the sharp drop in price was the overall increase in the supply of natural gas. More

and more natural gas was able to be recovered by fracturing shale deposits, a process called fracking. Fracking, which is controversial for environmental reasons, enabled the recovery of known reserves of natural gas that previously were not economically feasible to tap. Kinder Morgan's control of 80,000-plus miles of pipeline likely made moving the gas from wellheads to end users smoother and allowed for an even greater benefit from the increased supply.

Key Concepts and Summary

The U.S. economy experienced a wave of deregulation in the late 1970s and early 1980s, when a number of government regulations that had set prices and quantities produced in a number of industries were eliminated. Major accounting scandals in the early 2000s and, more recently, the Great Recession have spurred new regulation to prevent similar occurrences in the future. Regulatory capture occurs when the industries being regulated end up having a strong influence over what regulations exist.

Self-Check Questions

Use the following information to answer the next three questions. In the years before wireless phones, when telephone technology required having a wire running to every home, it seemed plausible that telephone service had diminishing average costs and might need to be regulated like a natural monopoly. For most of the twentieth century, the national U.S. phone company was AT&T, and the company functioned as a regulated monopoly. Think about the deregulation of the U.S. telecommunications industry that has happened over the last few decades. (This is not a research assignment, but a thought assignment based on what you have learned in this chapter.)

Exercise:

Problem: What real world changes made the deregulation possible?

Solution:

Improvements in technology that allowed phone calls to be made via microwave transmission, communications satellites, and other wireless technologies.

Exercise:

Problem: What are some of the benefits of the deregulation?

Solution:

More consumer choice. Cheaper phone calls, especially long distance. Better-quality phone service in many cases. Cheaper, faster, and better-quality data transmission. Spin-off technologies like free Internet-based calling and video calling.

Exercise:

Problem: What might some of the negatives of deregulation be?

Solution:

More choice can sometimes make for difficult decisions—not knowing if you got the best plan for your situation, for example. Some phone service providers are less reliable than AT&T used to be.

Review Questions

Exercise:

Problem:

What is deregulation? Name some industries that have been deregulated in the United States.

Exercise:

Problem: What is regulatory capture?

Exercise:

Problem:

Why does regulatory capture reduce the persuasiveness of the case for regulating industries for the benefit of consumers?

Critical Thinking Questions**Exercise:****Problem:**

Deregulation, like all changes in government policy, always has pluses and minuses. What do you think some of the minuses might be for airline deregulation?

Exercise:**Problem:**

Do you think it is possible for government to outlaw everything that businesses could do wrong? If so, why does government not do that? If not, how can regulation stay ahead of rogue businesses that push the limits of the system until it breaks?

References

Bishop, Todd. 2014. "Microsoft Exec Admits New Reality: Market Share No Longer 90% — It's 14%." *GeekWire*. Accessed March 27, 2015.
<http://www.geekwire.com/2014/microsoft-exec-admits-new-reality-market-share-longer-90-14/>.

Catan, T., & DeZemmer, R. "Kinder-El Paso Merger to Face Antitrust Scrutiny," *Wall Street Journal*. October 19, 2011.

Collins, A. "Tallgrass Energy to Acquire Kinder Morgan Assets for \$1.8B." *The Middle Market*, Accessed August 2013.

<http://www.themiddlemarket.com/news/tallgrass-energy-to-acquire-kinder-morgan-assets-for-1-point-8-billion-232862-1.html>.

Conoco Phillips. "Why Natural Gas." Accessed August 2013.
<http://www.powerincooperation.com/en/pages/useful.html?gclid=COXg7rH3hrGCFWlp7AodtkgA3g>.

De la Merced, M. (2012, August 20). "Kinder Morgan to Sell Assets to Tallgrass for \$1.8 Billion." *The New York Times*. August 20, 2012.

The Federal Trade Commission. n.d. "The Federal Trade Commission's (FTC) Mission." Accessed March 27, 2015.
<https://www.ftc.gov/system/files/documents/reports/2014-one-page-ftc-performance-snapshot/150218fy14snapshot.pdf>.

Kahn, C. "Kinder Morgan to Buy El Paso Corp. for \$20.7B." *USA Today*. October 17, 2011.

Kinder Morgan. (2013). "Investor Information." Accessed August 2013.
<http://www.kindermorgan.com/investor/>.

Rogowsky, Mark. 2014. "There'd Be No Wireless Wars Without The Blocked T-Mobile Merger, So Where Does That Leave Comcast-TWC?" *Forbes.com*. Accessed March 12, 2015.
<http://www.forbes.com/sites/markrogowsky/2014/08/27/t-mobile-and-sprint-continue-to-battle-thanks-to-the-government/>.

U.S. Energy Information Administration (a). "Natural Gas Consumption by End User." Accessed May 31, 2013.
http://www.eia.gov/dnav/ng/ng_cons_sum_dcu_nus_a.htm.

U.S. Energy Information Administration (b). "Natural Gas Prices." Accessed June 28, 2013.
http://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_a.htm.

The Wall Street Journal. 2015. "Auto Sales." Accessed April 10, 2015.
http://online.wsj.com/mdc/public/page/2_3022-autosales.html.

Glossary

regulatory capture

when the firms supposedly being regulated end up playing a large role in setting the regulations that they will follow and as a result, they “capture” the people doing the regulation, usually through the promise of a job in that “regulated” industry once their term in government has ended.

A World of Money

(Credit:
modification of
work by
epSos.de/Flickr
Creative
Commons)



Note:**More than Meets the Eye in the Congo**

How much do you interact with the global financial system? Do you think not much? Think again. Suppose you take out a student loan, or you deposit money into your bank account. You just affected domestic savings and borrowing. Now say you are at the mall and buy two T-shirts “made in China,” and later contribute to a charity that helps refugees. What is the impact? You affected how much money flows into and out of the United States. If you open an IRA savings account and put money in an international mutual fund, you are involved in the flow of money overseas. While your involvement may not seem as influential as someone like the president, who can increase or decrease foreign aid and, thereby, have a huge impact on money flows in and out of the country, you do interact with the global financial system on a daily basis.

The balance of payments—a term you will meet soon—seems like a huge topic, but once you learn the specific components of trade and money, it all makes sense. Along the way, you may have to give up some common misunderstandings about trade and answer some questions: If a country is running a trade deficit, is that bad? Is a trade surplus good? For example, look at the Democratic Republic of Congo (often referred to as “Congo”), a large country in Central Africa. In 2013, it ran a trade surplus of \$1 billion, so it must be doing well, right? In contrast, the trade deficit in the United States was \$508 billion in 2013. Do these figures suggest that the economy in the United States is doing worse than the Congolese economy? Not necessarily. The U.S. trade deficit tends to worsen as the economy strengthens. In contrast, high poverty rates in the Congo persist, and these rates are not going down even with the positive trade balance. Clearly, it is more complicated than simply asserting that running a trade deficit is bad for the economy. You will learn more about these issues and others in this chapter.

Note:**Introduction to International Trade and Capital Flows**

In this chapter, you will learn about:

- Measuring Trade Balances
- Trade Balances in Historical and International Context
- Trade Balances and Flows of Financial Capital
- The National Saving and Investment Identity
- The Pros and Cons of Trade Deficits and Surpluses
- The Difference between Level of Trade and the Trade Balance

The **balance of trade** (or trade balance) is any gap between a nation's dollar value of its exports, or what its producers sell abroad, and a nation's dollar worth of imports, or the foreign-made products and services that households and businesses purchase. Recall from [The Macroeconomic Perspective](#) that if exports exceed imports, the economy is said to have a trade surplus. If imports exceed exports, the economy is said to have a trade deficit. If exports and imports are equal, then trade is balanced. But what happens when trade is out of balance and large trade surpluses or deficits exist?

Germany, for example, has had substantial trade surpluses in recent decades, in which exports have greatly exceeded imports. According to the Central Intelligence Agency's *The World Factbook*, in 2013, Germany ran a trade surplus of \$260 billion. In contrast, the U.S. economy in recent decades has experienced large trade deficits, in which imports have considerably exceeded exports. In 2014, for example, U.S. imports exceeded exports by \$539 billion.

A series of financial crises triggered by unbalanced trade can lead economies into deep recessions. These crises begin with large trade deficits. At some point, foreign investors become pessimistic about the economy and move their money to other countries. The economy then drops into deep recession, with real GDP often falling up to 10% or more in a single year. This happened to Mexico in 1995 when their GDP fell 8.1%. A number of countries in East Asia—Thailand, South Korea, Malaysia, and Indonesia—came down with the same economic illness in 1997–1998 (called the Asian Financial Crisis). In the late 1990s and into the early 2000s, Russia and Argentina had the identical experience. What are the connections between

imbalances of trade in goods and services and the flows of international financial capital that set off these economic avalanches?

We will start by examining the balance of trade in more detail, by looking at some patterns of trade balances in the United States and around the world. Then we will examine the intimate connection between international flows of goods and services and international flows of financial capital, which to economists are really just two sides of the same coin. It is often assumed that trade surpluses like those in Germany must be a positive sign for an economy, while trade deficits like those in the United States must be harmful. As it turns out, both trade surpluses and deficits can be either good or bad. We will see why in this chapter.

Measuring Trade Balances

By the end of this section, you will be able to:

- Explain merchandise trade balance, current account balance, and unilateral transfers
- Identify components of the U.S. current account balance
- Calculate the merchandise trade balance and current account balance using import and export data for a country

A few decades ago, it was common to track the solid or physical items that were transported by planes, trains, and trucks between countries as a way of measuring the balance of trade. This measurement is called the **merchandise trade balance**. In most high-income economies, including the United States, goods make up less than half of a country's total production, while services compose more than half. The last two decades have seen a surge in international trade in services, powered by technological advances in telecommunications and computers that have made it possible to export or import customer services, finance, law, advertising, management consulting, software, construction engineering, and product design. Most global trade still takes the form of goods rather than services, and the merchandise trade balance is still announced by the government and reported prominently in the newspapers. Old habits are hard to break. Economists, however, typically rely on broader measures such as the balance of trade or the **current account balance** which includes other international flows of income and foreign aid.

Components of the U.S. Current Account Balance

[\[link\]](#) breaks down the four main components of the U.S. current account balance for the last quarter of 2014 (seasonally adjusted). The first line shows the merchandise trade balance; that is, exports and imports of goods. Because imports exceed exports, the trade balance in the final column is negative, showing a merchandise trade deficit.

	Value of Exports (money flowing into the United States)	Value of Imports (money flowing out of the United States)	Balance
Goods	\$410.0	\$595.5	−\$185.3
Services	\$180.4	\$122.3	\$58.1
Income receipts and payments	\$203.0	\$152.4	\$50.6
Unilateral transfers	\$27.3	\$64.4	−\$37.1
Current account balance	\$820.7	\$934.4	−\$113.7

Components of the U.S. Current Account Balance for 2014 (in billions)

The second row of [\[link\]](#) provides data on trade in services. Here, the U.S. economy is running a surplus. Although the level of trade in services is still relatively small compared to trade in goods, the importance of services has expanded substantially over the last few decades. For example, U.S. exports of services were equal to about one-half of U.S. exports of goods in 2014, compared to one-fifth in 1980.

The third component of the current account balance, labeled “income payments,” refers to money received by U.S. financial investors on their foreign investments (money flowing into the United States) and payments to foreign investors who had invested their funds here (money flowing out of the United States). The reason for including this money on foreign investment in the overall measure of trade, along with goods and services, is that, from an economic perspective, income is just as much an economic

transaction as shipments of cars or wheat or oil: it is just trade that is happening in the financial capital market.

The final category of the current account balance is **unilateral transfers**, which are payments made by government, private charities, or individuals in which money is sent abroad without any direct good or service being received. Economic or military assistance from the U.S. government to other countries fits into this category, as does spending abroad by charities to address poverty or social inequalities. When an individual in the United States sends money overseas, it is also counted in this category. The current account balance treats these unilateral payments like imports, because they also involve a stream of payments leaving the country. For the U.S. economy, unilateral transfers are almost always negative. This pattern, however, does not always hold. In 1991, for example, when the United States led an international coalition against Saddam Hussein's Iraq in the Gulf War, many other nations agreed that they would make payments to the United States to offset the U.S. war expenses. These payments were large enough that, in 1991, the overall U.S. balance on unilateral transfers was a positive \$10 billion.

Key Concepts and Summary

The trade balance measures the gap between a country's exports and its imports. In most high-income economies, goods make up less than half of a country's total production, while services compose more than half. The last two decades have seen a surge in international trade in services; however, most global trade still takes the form of goods rather than services. The current account balance includes the trade in goods, services, and money flowing into and out of a country from investments and unilateral transfers.

Self-Check Questions

Exercise:

Problem:

If foreign investors buy more U.S. stocks and bonds, how would that show up in the current account balance?

Solution:

The stock and bond values will not show up in the current account. However, the dividends from the stocks and the interest from the bonds show up as an import to income in the current account.

Exercise:**Problem:**

If the trade deficit of the United States increases, how is the current account balance affected?

Solution:

It becomes more negative as imports, which are a negative to the current account, are growing faster than exports, which are a positive.

Exercise:**Problem:**

State whether each of the following events involves a financial flow to the Mexican economy or a financial flow out of the Mexican economy:

- a. Mexico imports services from Japan
- b. Mexico exports goods to Canada
- c. U.S. investors receive a return from past financial investments in Mexico

Solution:

- a. Money flows out of the Mexican economy.
- b. Money flows into the Mexican economy.

c. Money flows out of the Mexican economy.

Review Questions

Exercise:

Problem:

If imports exceed exports, is it a trade deficit or a trade surplus? What about if exports exceed imports?

Exercise:

Problem: What is included in the current account balance?

Critical Thinking Questions

Exercise:

Problem:

From time to time, a government official will argue that a country should strive for both a trade surplus and a healthy inflow of capital from abroad. Explain why such a statement is economically impossible.

Exercise:

Problem:

A government official announces a new policy. The country wishes to eliminate its trade deficit, but will strongly encourage financial investment from foreign firms. Explain why such a statement is contradictory.

Problems

Exercise:

Problem:

In 2001, the economy of the United Kingdom exported goods worth £192 billion and services worth another £77 billion. It imported goods worth £225 billion and services worth £66 billion. Receipts of income from abroad were £140 billion while income payments going abroad were £131 billion. Government transfers from the United Kingdom to the rest of the world were £23 billion, while various U.K government agencies received payments of £16 billion from the rest of the world.

- a. Calculate the U.K. merchandise trade deficit for 2001.
- b. Calculate the current account balance for 2001.
- c. Explain how you decided whether payments on foreign investment and government transfers counted on the positive or the negative side of the current account balance for the United Kingdom in 2001.

References

Central Intelligence Agency. "The World Factbook." Last modified October 31, 2013. <https://www.cia.gov/library/publications/the-world-factbook/geos/gm.html>.

U.S. Department of Commerce. "Bureau of Economic Analysis." Last modified December 1, 2013. <http://www.bea.gov/>.

U.S. Department of Commerce. "United States Census Bureau." <http://www.census.gov/>.

Glossary

balance of trade (trade balance)

the gap, if any, between a nation's exports and imports

current account balance

a broad measure of the balance of trade that includes trade in goods and services, as well as international flows of income and foreign aid

merchandise trade balance

the balance of trade looking only at goods

unilateral transfers

“one-way payments” made by governments, private entities, or individuals that are sent abroad with nothing received in return

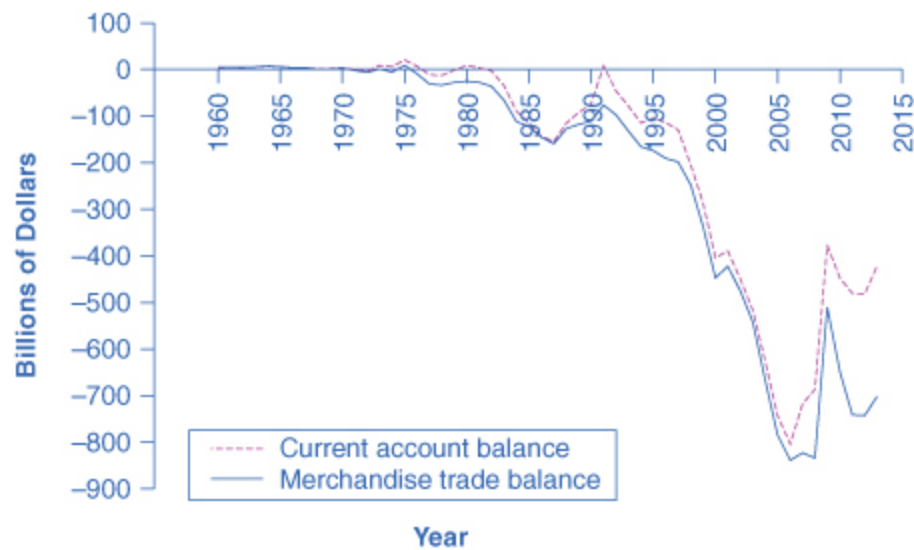
Trade Balances in Historical and International Context

By the end of this section, you will be able to:

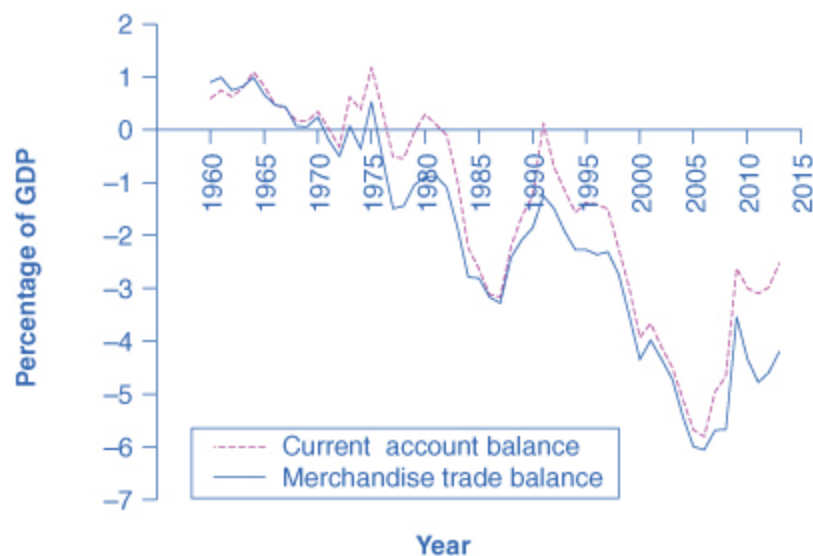
- Analyze graphs of the current account balance and the merchandise trade balance
- Identify patterns in U.S. trade surpluses and deficits
- Compare the U.S. trade surpluses and deficits to other countries' trade surpluses and deficits

The history of the U.S. current account balance in recent decades is presented in several different ways. [\[link\]](#) (a) shows the current account balance and the merchandise trade balance in dollar terms. [\[link\]](#) (b) shows the current account balance and merchandise account balance yet again, this time presented as a share of the GDP for that year. By dividing the trade deficit in each year by GDP in that year, [\[link\]](#) (b) factors out both inflation and growth in the real economy.

Current Account Balance and Merchandise Trade Balance, 1960–2013



(a) The current account and merchandise trade balance in nominal dollars



(b) The current account and merchandise trade balance as a percentage of GDP

(a) The current account balance and the merchandise trade balance in billions of dollars from 1960 to 2013. If the lines are above zero dollars, the United States was running a positive trade balance and current account balance. If the lines fall below zero dollars, the United States is

running a trade deficit and a deficit in its current account balance. (b) These same items—trade balance and current account balance—are shown in relationship to the size of the U.S. economy, or GDP, from 1960 to 2012.

By either measure, the general pattern of the U.S. balance of trade is clear. From the 1960s into the 1970s, the U.S. economy had mostly small trade surpluses—that is, the graphs of [\[link\]](#) show positive numbers. However, starting in the 1980s, the trade deficit increased rapidly, and after a tiny surplus in 1991, the current account trade deficit got even larger in the late 1990s and into the mid-2000s. However, the trade deficit declined in 2009 after the recession had taken hold.

[\[link\]](#) shows the U.S. trade picture in 2013 compared with some other economies from around the world. While the U.S. economy has consistently run trade deficits in recent years, Japan and many European nations, among them France and Germany, have consistently run trade surpluses. Some of the other countries listed include Brazil, the largest economy in Latin America; Nigeria, the largest economy in Africa; and China, India, and Korea. The first column offers one measure of the globalization of an economy: **exports of goods and services as a percentage of GDP**. The second column shows the trade balance. Most of the time, most countries have trade surpluses or deficits that are less than 5% of GDP. As you can see, the U.S. current account balance is –2.3% of GDP, while Germany's is 7.4% of GDP.

	Exports of Goods and Services	Current Account Balance
--	-------------------------------	-------------------------

	Exports of Goods and Services	Current Account Balance
United States	13.5%	−2.3%
Japan	16.2%	0.7%
Germany	45.6%	7.4%
United Kingdom	29.8%	−4.2%
Canada	30.1%	−3.2%
Sweden	43.8%	6.7%
Korea	53.9%	5.4%
Mexico	31.7%	−2.3%
Brazil	12.6%	−3.6%
China	26.4%	2.0%
India	24.8%	−2.6%
Nigeria	18.0%	4.1%
World	-	0.0%

Level and Balance of Trade in 2013 (figures as a percentage of GDP,
Source: <http://data.worldbank.org/indicator/BN.CAB.XOKA.GD.ZS>)

Key Concepts and Summary

The United States developed large trade surpluses in the early 1980s, swung back to a tiny trade surplus in 1991, and then had even larger trade deficits in the late 1990s and early 2000s. As we will see below, a trade deficit necessarily means a net inflow of financial capital from abroad, while a trade surplus necessarily means a net outflow of financial capital from an economy to other countries.

Self-Check Questions

Exercise:

Problem:

In what way does comparing a country's exports to GDP reflect how globalized it is?

Solution:

GDP is a dollar value of all production of goods and services. Exports are produced domestically but shipped abroad. The percent ratio of exports to GDP gives us an idea of how important exports are to the national economy out of all goods and services produced. For example, exports represent only 14% of U.S. GDP, but 50% of Germany's GDP

Exercise:

Problem:

Canada's GDP is \$1,800 billion and its exports are \$542 billion. What is Canada's export ratio?

Solution:

Divide \$542 billion by \$1,800 billion.

Exercise:

Problem:

The GDP for the United States is \$16,800 billion and its current account balance is −\$400 billion. What percent of GDP is the current account balance?

Solution:

Divide −\$400 billion by \$16,800 billion.

Exercise:**Problem:**

Why does the trade balance and the current account balance track so closely together over time?

Solution:

The trade balance is the difference between exports and imports. The current account balance includes this number (whether it is a trade balance or a trade surplus), but also includes international flows of money from global investments.

Review Question**Exercise:****Problem:**

In recent decades, has the U.S. trade balance usually been in deficit, surplus, or balanced?

Critical Thinking Questions**Exercise:**

Problem:

If a country is a big exporter, is it more exposed to global financial crises?

Exercise:**Problem:**

If countries reduced trade barriers, would the international flows of money increase?

Glossary

exports of goods and services as a percentage of GDP

the dollar value of exports divided by the dollar value of a country's GDP

Trade Balances and Flows of Financial Capital

By the end of this section, you will be able to:

- Explain the connection between trade balances and flows of financial capital
- Calculate comparative advantage
- Explain balanced trade in terms of investment and capital flows

As economists see it, trade surpluses can be either good or bad, depending on circumstances, and trade deficits can be good or bad, too. The challenge is to understand how the international flows of goods and services are connected with international flows of **financial capital**. In this module we will illustrate the intimate connection between trade balances and flows of financial capital in two ways: a parable of trade between Robinson Crusoe and Friday, and a circular flow diagram representing flows of trade and payments.

A Two-Person Economy: Robinson Crusoe and Friday

To understand how economists view trade deficits and surpluses, consider a parable based on the story of Robinson Crusoe. Crusoe, as you may remember from the classic novel by Daniel Defoe first published in 1719, was shipwrecked on a desert island. After living alone for some time, he is joined by a second person, whom he names Friday. Think about the balance of trade in a two-person economy like that of Robinson and Friday.

Robinson and Friday trade goods and services. Perhaps Robinson catches fish and trades them to Friday for coconuts. Or Friday weaves a hat out of tree fronds and trades it to Robinson for help in carrying water. For a period of time, each individual trade is self-contained and complete. Because each trade is voluntary, both Robinson and Friday must feel that they are receiving fair value for what they are giving. As a result, each person's exports are always equal to his imports, and trade is always in balance between the two. Neither person experiences either a trade deficit or a trade surplus.

However, one day Robinson approaches Friday with a proposition. Robinson wants to dig ditches for an irrigation system for his garden, but he knows that if he starts this project, he will not have much time left to fish and gather coconuts to feed himself each day. He proposes that Friday supply him with a certain number of fish and coconuts for several months, and then after that time, he promises to repay Friday out of the extra produce that he will be able to grow in his irrigated garden. If Friday accepts this offer, then a trade imbalance comes into being. For several months, Friday will have a trade surplus: that is, he is exporting to Robinson more than he is importing. More precisely, he is giving Robinson fish and coconuts, and at least for the moment, he is receiving nothing in return. Conversely, Robinson will have a trade deficit, because he is importing more from Friday than he is exporting.

This parable raises several useful issues in thinking about what a trade deficit and a trade surplus really mean in economic terms. The first issue raised by this story of Robinson and Friday is this: Is it better to have a trade surplus or a trade deficit? The answer, as in any voluntary market interaction, is that if both parties agree to the transaction, then they may both be better off. Over time, if Robinson's irrigated garden is a success, it is certainly possible that both Robinson and Friday can benefit from this agreement.

A second issue raised by the parable: What can go wrong? Robinson's proposal to Friday introduces an element of uncertainty. Friday is, in effect, making a loan of fish and coconuts to Robinson, and Friday's happiness with this arrangement will depend on whether that loan is repaid as planned, in full and on time. Perhaps Robinson spends several months loafing and never builds the irrigation system. Or perhaps Robinson has been too optimistic about how much he will be able to grow with the new irrigation system, which turns out not to be very productive. Perhaps, after building the irrigation system, Robinson decides that he does not want to repay Friday as much as previously agreed. Any of these developments will prompt a new round of negotiations between Friday and Robinson. Friday's attitude toward these renegotiations is likely to be shaped by why the repayment failed. If Robinson worked very hard and the irrigation system just did not increase production as intended, Friday may have some

sympathy. If Robinson loafed or if he just refuses to pay, Friday may become irritated.

A third issue raised by the parable of Robinson and Friday is that an intimate relationship exists between a trade deficit and international borrowing, and between a trade surplus and international lending. The size of Friday’s trade surplus is exactly how much he is lending to Robinson. The size of Robinson’s trade deficit is exactly how much he is borrowing from Friday. Indeed, to economists, a trade surplus literally means the same thing as an outflow of financial capital, and a trade deficit literally means the same thing as an inflow of financial capital. This last insight is worth exploring in greater detail, which we will do in the following section.

The story of Robinson and Friday also provides a good opportunity to consider the law of comparative advantage, which you learn more about in the [International Trade](#) chapter. The following Work It Out feature steps you through calculating comparative advantage for the wheat and cloth traded between the United States and Great Britain in the 1800s.

Note: Calculating Comparative Advantage In the 1800s, the United States and Britain traded wheat and cloth. [link] shows the varying hours of labor per unit of output.				
	Wheat (in bushels)	Cloth (in yards)	Relative labor cost of wheat (P_w/P_c)	Relative labor cost of cloth (P_c/P_w)

	Wheat (in bushels)	Cloth (in yards)	Relative labor cost of wheat (P_w/P_c)	Relative labor cost of cloth (P_c/P_w)
United States	8	9	8/9	9/8
Britain	4	3	4/3	3/4

Step 1. Observe from [\[link\]](#) that, in the United States, it takes eight hours to supply a bushel of wheat and nine hours to supply a yard of cloth. In contrast, it takes four hours to supply a bushel of wheat and three hours to supply a yard of cloth in Britain.

Step 2. Recognize the difference between absolute advantage and comparative advantage. Britain has an absolute advantage (lowest cost) in each good, since it takes a lower amount of labor to make each good in Britain. Britain also has a comparative advantage in the production of cloth (lower opportunity cost in cloth ($3/4$ versus $9/8$)). The United States has a comparative advantage in wheat production (lower opportunity cost of $8/9$ versus $4/3$).

Step 3. Determine the relative price of one good in terms of the other good. The price of wheat, in this example, is the amount of cloth you have to give up. To find this price, convert the hours per unit of wheat and cloth into units per hour. To do so, observe that in the United States it takes eight hours to make a bushel of wheat, so $1/8$ of a bushel of wheat can be made in an hour. It takes nine hours to make a yard of cloth in the United States, so $1/9$ of a yard of cloth can be made in an hour. If you divide the amount of cloth ($1/9$ of a yard) by the amount of wheat you give up ($1/8$ of a bushel) in an hour, you find the price ($8/9$) of one good (wheat) in terms of the other (cloth).

The Balance of Trade as the Balance of Payments

The connection between trade balances and international flows of financial capital is so close that the balance of trade is sometimes described as the balance of payments. Each category of the current account balance involves a corresponding flow of payments between a given country and the rest of the world economy.

A current account deficit means that, the country is a net borrower from abroad. Conversely, a positive current account balance means a country is a net lender to the rest of the world. Just like the parable of Robinson and Friday, the lesson is that a trade surplus means an overall outflow of financial investment capital, as domestic investors put their funds abroad, while the deficit in the current account balance is exactly equal to the overall or net inflow of foreign investment capital from abroad.

It is important to recognize that an inflow and outflow of foreign capital does not necessarily refer to a debt that governments owe to other governments, although government debt may be part of the picture. Instead, these international flows of financial capital refer to all of the ways in which private investors in one country may invest in another country—by buying real estate, companies, and financial investments like stocks and bonds.

Key Concepts and Summary

International flows of goods and services are closely connected to the international flows of financial capital. A current account deficit means that, after taking all the flows of payments from goods, services, and income together, the country is a net borrower from the rest of the world. A current account surplus is the opposite and means the country is a net lender to the rest of the world.

Self-Check Questions

Exercise:

Problem:

State whether each of the following events involves a financial flow to the U.S. economy or away from the U.S. economy:

- a. Export sales to Germany
 - b. Returns being paid on past U.S. financial investments in Brazil
 - c. Foreign aid from the U.S. government to Egypt
 - d. Imported oil from the Russian Federation
 - e. Japanese investors buying U.S. real estate
-

Solution:

- a. An export sale to Germany involves a financial flow from Germany to the U.S. economy.
- b. The issue here is not U.S. investments in Brazil, but the return paid on those investments, which involves a financial flow from the Brazilian economy to the U.S. economy.
- c. Foreign aid from the United States to Egypt is a financial flow from the United States to Egypt.
- d. Importing oil from the Russian Federation means a flow of financial payments from the U.S. economy to the Russian Federation.
- e. Japanese investors buying U.S. real estate is a financial flow from Japan to the U.S. economy.

Exercise:**Problem:**

How does the bottom portion of [\[link\]](#), showing the international flow of investments and capital, differ from the upper portion?

Solution:

The top portion tracks the flow of exports and imports and the payments for those. The bottom portion is looking at international

financial investments and the outflow and inflow of monies from those investments. These investments can include investments in stocks and bonds or real estate abroad, as well as international borrowing and lending.

Exercise:

Problem:

Explain the relationship between a current account deficit or surplus and the flow of funds.

Solution:

If more monies are flowing out of the country (for example, to pay for imports) it will make the current account more negative or less positive, and if more monies are flowing into the country, it will make the current account less negative or more positive.

Review Question

Exercise:

Problem:

Does a trade surplus mean an overall inflow of financial capital to an economy, or an overall outflow of financial capital? What about a trade deficit?

Critical Thinking Question

Exercise:

Problem:

Is it better for your country to be an international lender or borrower?

Glossary

financial capital

the international flows of money that facilitates trade and investment

The Pros and Cons of Trade Deficits and Surpluses

By the end of this section, you will be able to:

- Identify three ways in which borrowing money or running a trade deficit can result in a healthy economy
- Identify three ways in which borrowing money or running a trade deficit can result in a weaker economy

Because flows of trade always involve flows of financial payments, flows of international trade are actually the same as flows of international financial capital. The question of whether trade deficits or surpluses are good or bad for an economy is, in economic terms, exactly the same question as whether it is a good idea for an economy to rely on net inflows of financial capital from abroad or to make net investments of financial capital abroad. Conventional wisdom often holds that borrowing money is foolhardy, and that a prudent country, like a prudent person, should always rely on its own resources. While it is certainly possible to borrow too much—as anyone with an overloaded credit card can testify—borrowing at certain times can also make sound economic sense. For both individuals and countries, there is no economic merit in a policy of abstaining from participation in financial capital markets.

It makes economic sense to borrow when you are buying something with a long-run payoff; that is, when you are making an investment. For this reason, it can make economic sense to borrow for a college education, because the education will typically allow you to earn higher wages, and so to repay the loan and still come out ahead. It can also make sense for a business to borrow in order to purchase a machine that will last 10 years, as long as the machine will increase output and profits by more than enough to repay the loan. Similarly, it can make economic sense for a national economy to borrow from abroad, as long as the money is wisely invested in ways that will tend to raise the nation's economic growth over time. Then, it will be possible for the national economy to repay the borrowed money over time and still end up better off than before.

One vivid example of a country that borrowed heavily from abroad, invested wisely, and did perfectly well is the United States during the

nineteenth century. The United States ran a trade deficit in 40 of the 45 years from 1831 to 1875, which meant that it was importing capital from abroad over that time. However, that financial capital was, by and large, invested in projects like railroads that brought a substantial economic payoff. (See the following Clear It Up feature for more on this.)

A more recent example along these lines is the experience of South Korea, which had trade deficits during much of the 1970s—and so was an importer of capital over that time. However, South Korea also had high rates of investment in physical plant and equipment, and its economy grew rapidly. From the mid-1980s into the mid-1990s, South Korea often had trade surpluses—that is, it was repaying its past borrowing by sending capital abroad.

In contrast, some countries have run large trade deficits, borrowed heavily in global capital markets, and ended up in all kinds of trouble. Two specific sorts of trouble are worth examining. First, a borrower nation can find itself in a bind if the incoming funds from abroad are not invested in a way that leads to increased productivity. Several of the large economies of Latin America, including Mexico and Brazil, ran large trade deficits and borrowed heavily from abroad in the 1970s, but the inflow of financial capital did not boost productivity sufficiently, which meant that these countries faced enormous troubles repaying the money borrowed when economic conditions shifted during the 1980s. Similarly, it appears that a number of African nations that borrowed foreign funds in the 1970s and 1980s did not invest in productive economic assets. As a result, several of those countries later faced large interest payments, with no economic growth to show for the borrowed funds.

Note:

Are trade deficits always harmful?

For most years of the nineteenth century, U.S. imports exceeded exports and the U.S. economy had a trade deficit. Yet the string of trade deficits did not hold back the economy at all; instead, the trade deficits contributed to the strong economic growth that gave the U.S. economy the highest per capita GDP in the world by around 1900.

The U.S. trade deficits meant that the U.S. economy was receiving a net inflow of foreign capital from abroad. Much of that foreign capital flowed into two areas of investment—railroads and public infrastructure like roads, water systems, and schools—which were important to helping the growth of the U.S. economy.

The effect of foreign investment capital on U.S. economic growth should not be overstated. In most years the foreign financial capital represented no more than 6–10% of the funds used for overall physical investment in the economy. Nonetheless, the trade deficit and the accompanying investment funds from abroad were clearly a help, not a hindrance, to the U.S. economy in the nineteenth century.

A second “trouble” is: What happens if the foreign money flows in, and then suddenly flows out again? This scenario was raised at the start of the chapter. In the mid-1990s, a number of countries in East Asia—Thailand, Indonesia, Malaysia, and South Korea—ran large trade deficits and imported capital from abroad. However, in 1997 and 1998 many foreign investors became concerned about the health of these economies, and quickly pulled their money out of stock and bond markets, real estate, and banks. The extremely rapid departure of that foreign capital staggered the banking systems and economies of these countries, plunging them into deep recession. We investigate and discuss the links between international capital flows, banks, and recession in [The Impacts of Government Borrowing](#).

While a trade deficit is not always harmful, there is no guarantee that running a trade surplus will bring robust economic health. For example, Germany and Japan ran substantial trade surpluses for most of the last three decades. Regardless of their persistent trade surpluses, both countries have experienced occasional recessions and neither country has had especially robust annual growth in recent years. Read more about Japan’s trade surplus in the next Clear It Up feature.

The sheer size and persistence of the U.S. trade deficits and inflows of foreign capital since the 1980s are a legitimate cause for concern. The huge U.S. economy will not be destabilized by an outflow of international capital as easily as, say, the comparatively tiny economies of Thailand and

Indonesia were in 1997–1998. Even an economy that is not knocked down, however, can still be shaken. American policymakers should certainly be paying attention to those cases where a pattern of extensive and sustained current account deficits and foreign borrowing has gone badly—if only as a cautionary tale.

Note:

Are trade surpluses always beneficial? Considering Japan since the 1990s. Perhaps no economy around the world is better known for its trade surpluses than Japan. Since 1990, the size of these surpluses has often been near \$100 billion per year. When Japan's economy was growing vigorously in the 1960s and 1970s, its large trade surpluses were often described, especially by non-economists, as either a cause or a result of its robust economic health. But from a standpoint of economic growth, Japan's economy has been teetering in and out of recession since 1990, with real GDP growth averaging only about 1% per year, and an unemployment rate that has been creeping higher. Clearly, a whopping trade surplus is no guarantee of economic good health.

Instead, Japan's trade surplus reflects that Japan has a very high rate of domestic savings, more than the Japanese economy can invest domestically, and so the extra funds are invested abroad. In Japan's slow economy, the growth of consumption is relatively low, which also means that consumption of imports is relatively low. Thus, Japan's exports continually exceed its imports, leaving the trade surplus continually high. Recently, Japan's trade surpluses began to deteriorate. In 2013, Japan ran a trade deficit due to the high cost of imported oil.

Key Concepts and Summary

Trade surpluses are no guarantee of economic health, and trade deficits are no guarantee of economic weakness. Either trade deficits or trade surpluses can work out well or poorly, depending on whether the corresponding flows of financial capital are wisely invested.

Self-Check Questions

Exercise:

Problem:

For each of the following, indicate which type of government spending would justify a budget deficit and which would not.

- a. Increased federal spending on Medicare
- b. Increased spending on education
- c. Increased spending on the space program
- d. Increased spending on airports and air traffic control

Solution:

- a. Increased federal spending on Medicare may not increase productivity, so a budget deficit is not justified.
- b. Increased spending on education will increase productivity and foster greater economic growth, so a budget deficit is justified.
- c. Increased spending on the space program may not increase productivity, so a budget deficit is not justified.
- d. Increased spending on airports and air traffic control will increase productivity and foster greater economic growth, so a budget deficit is justified.

Exercise:

Problem:

How did large trade deficits hurt the East Asian countries in the mid 1980's? (Recall that trade deficits are equivalent to inflows of financial capital from abroad.)

Solution:

Foreign investors worried about repayment so they began to pull money out of these countries. The money can be pulled out of stock

and bond markets, real estate, and banks.

Exercise:

Problem:

Describe a scenario in which a trade surplus benefits an economy and one in which a trade surplus is occurring in an economy that performs poorly. What key factor or factors are making the difference in the outcome that results from a trade surplus?

Solution:

A rapidly growing trade surplus could result from a number of factors, so you would not want to be too quick to assume a specific cause. However, if the choice is between whether the economy is in recession or growing rapidly, the answer would have to be recession. In a recession, demand for all goods, including imports, has declined; however, demand for exports from other countries has not necessarily altered much, so the result is a larger trade surplus.

Review Questions

Exercise:

Problem:

When is a trade deficit likely to work out well for an economy? When is it likely to work out poorly?

Exercise:

Problem:

Does a trade surplus help to guarantee strong economic growth?

Critical Thinking Question

Exercise:

Problem:

What is more important, a country's current account balance or the growth of GDP? Why?

References

Tabuchi, Hiroko. "Japan Reports a \$78 Billion Trade Deficit for 2012." *The New York Times*, January 23, 2013.

http://www.nytimes.com/2013/01/24/business/global/japan-reports-a-78-billion-trade-deficit-for-2012.html?_r=0.

World Bank. 2014. "Exports of Goods and Services (% of GDP)." Accessed April 13, 2015.

<http://data.worldbank.org/indicator/NE.EXP.GNFS.ZS/countries>.

The Difference between Level of Trade and the Trade Balance

By the end of this section, you will be able to:

- Identify three factors that influence a country's level of trade
- Differentiate between balance of trade and level of trade

A nation's *level* of trade may at first sound like much the same issue as the *balance* of trade, but these two are actually quite separate. It is perfectly possible for a country to have a very high level of trade—measured by its exports of goods and services as a share of its GDP—while it also has a near-balance between exports and imports. A high level of trade indicates that a good portion of the nation's production is exported. It is also possible for a country's trade to be a relatively low share of GDP, relative to global averages, but for the imbalance between its exports and its imports to be quite large. This general theme was emphasized earlier in [Measuring Trade Balances](#), which offered some illustrative figures on trade levels and balances.

A country's level of trade tells how much of its production it exports. This is measured by the percent of exports out of GDP. It indicates how globalized an economy is. Some countries, such as Germany, have a high level of trade—they export almost 50% of their total production. The balance of trade tells us if the country is running a trade surplus or trade deficit. A country can have a low level of trade but a high trade deficit. (For example, the United States only exports 14% of GDP, but it has a trade deficit of \$540 billion.)

Three factors strongly influence a nation's level of trade: the size of its economy, its geographic location, and its history of trade. Large economies like the United States can do much of their trading internally, while small economies like Sweden have less ability to provide what they want internally and tend to have higher ratios of exports and imports to GDP. Nations that are neighbors tend to trade more, since costs of transportation and communication are lower. Moreover, some nations have long and established patterns of international trade, while others do not.

Consequently, a relatively small economy like Sweden, with many nearby trading partners across Europe and a long history of foreign trade, has a high level of trade. Brazil and India, which are fairly large economies that have often sought to inhibit trade in recent decades, have lower levels of trade. Whereas, the United States and Japan are extremely large economies that have comparatively few nearby trading partners. Both countries actually have quite low levels of trade by world standards. The ratio of exports to GDP in either the United States or in Japan is about half of the world average.

The balance of trade is a separate issue from the level of trade. The United States has a low level of trade, but had enormous trade deficits for most years from the mid-1980s into the 2000s. Japan has a low level of trade by world standards, but has typically shown large trade surpluses in recent decades. Nations like Germany and the United Kingdom have medium to high levels of trade by world standards, but Germany had a moderate trade surplus in 2008, while the United Kingdom had a moderate trade deficit. Their trade picture was roughly in balance in the late 1990s. Sweden had a high level of trade and a large trade surplus in 2007, while Mexico had a high level of trade and a moderate trade deficit that same year.

In short, it is quite possible for nations with a relatively low level of trade, expressed as a percentage of GDP, to have relatively large trade deficits. It is also quite possible for nations with a near balance between exports and imports to worry about the consequences of high levels of trade for the economy. It is not inconsistent to believe that a high level of trade is potentially beneficial to an economy, because of the way it allows nations to play to their comparative advantages, and to also be concerned about any macroeconomic instability caused by a long-term pattern of large trade deficits. The following Clear It Up feature discusses how this sort of dynamic played out in Colonial India.

Note:

Are trade surpluses always beneficial? Considering Colonial India. India was formally under British rule from 1858 to 1947. During that time, India consistently had trade surpluses with Great Britain. Anyone who

believes that trade surpluses are a sign of economic strength and dominance while trade deficits are a sign of economic weakness must find this pattern odd, since it would mean that colonial India was successfully dominating and exploiting Great Britain for almost a century—which was not true.

Instead, India's trade surpluses with Great Britain meant that each year there was an overall flow of financial capital from India to Great Britain. In India, this flow of financial capital was heavily criticized as the “drain,” and eliminating the drain of financial capital was viewed as one of the many reasons why India would benefit from achieving independence.

Final Thoughts about Trade Balances

Trade deficits can be a good or a bad sign for an economy, and trade surpluses can be a good or a bad sign. Even a trade balance of zero—which just means that a nation is neither a net borrower nor lender in the international economy—can be either a good or bad sign. The fundamental economic question is not whether a nation's economy is borrowing or lending at all, but whether the particular borrowing or lending in the particular economic conditions of that country makes sense.

It is interesting to reflect on how public attitudes toward trade deficits and surpluses might change if we could somehow change the labels that people and the news media affix to them. If a trade deficit was called “attracting foreign financial capital”—which accurately describes what a trade deficit means—then trade deficits might look more attractive. Conversely, if a trade surplus were called “shipping financial capital abroad”—which accurately captures what a trade surplus does—then trade surpluses might look less attractive. Either way, the key to understanding trade balances is to understand the relationships between flows of trade and flows of international payments, and what these relationships imply about the causes, benefits, and risks of different kinds of trade balances. The first step along this journey of understanding is to move beyond knee-jerk reactions to terms like “trade surplus,” “trade balance,” and “trade deficit.”

Note:**More than Meets the Eye in the Congo**

Now that you see the big picture, you undoubtedly realize that all of the economic choices you make, such as depositing savings or investing in an international mutual fund, do influence the flow of goods and services as well as the flows of money around the world.

You now know that a trade surplus does not necessarily tell us whether an economy is doing well or not. The Democratic Republic of Congo ran a trade surplus in 2013, as we learned in the beginning of the chapter. Yet its current account balance was $-\$2.8$ billion. However, the return of political stability and the rebuilding in the aftermath of the civil war there has meant a flow of investment and financial capital into the country. In this case, a negative current account balance means the country is being rebuilt—and that is a good thing.

Key Concepts and Summary

There is a difference between the level of a country's trade and the balance of trade. The level of trade is measured by the percentage of exports out of GDP, or the size of the economy. Small economies that have nearby trading partners and a history of international trade will tend to have higher levels of trade. Larger economies with few nearby trading partners and a limited history of international trade will tend to have lower levels of trade. The level of trade is different from the trade balance. The level of trade depends on a country's history of trade, its geography, and the size of its economy. A country's balance of trade is the dollar difference between its exports and imports.

Trade deficits and trade surpluses are not necessarily good or bad—it depends on the circumstances. Even if a country is borrowing, if that money is invested in productivity-boosting investments it can lead to an improvement in long-term economic growth.

Self-Check Questions

Exercise:**Problem:**

The United States exports 14% of GDP while Germany exports about 50% of its GDP. Explain what that means.

Solution:

Germany has a higher level of trade than the United States. The United States has a large domestic economy so it has a large volume of internal trade.

Exercise:**Problem:**

Explain briefly whether each of the following would be more likely to lead to a higher level of trade for an economy, or a greater imbalance of trade for an economy.

- a. Living in an especially large country
 - b. Having a domestic investment rate much higher than the domestic savings rate
 - c. Having many other large economies geographically nearby
 - d. Having an especially large budget deficit
 - e. Having countries with a tradition of strong protectionist legislation shutting out imports
-

Solution:

- a. A large economy tends to have lower levels of international trade, because it can do more of its trade internally, but this has little impact on its trade imbalance.
- b. An imbalance between domestic physical investment and domestic saving (including government and private saving) will always lead to a trade imbalance, but has little to do with the level of trade.

- c. Many large trading partners nearby geographically increases the level of trade, but has little impact one way or the other on a trade imbalance.
- d. The answer here is not obvious. An especially large budget deficit means a large demand for financial capital which, according to the national saving and investment identity, makes it somewhat more likely that there will be a need for an inflow of foreign capital, which means a trade deficit.
- e. A strong tradition of discouraging trade certainly reduces the level of trade. However, it does not necessarily say much about the balance of trade, since this is determined by both imports and exports, and by national levels of physical investment and savings.

Review Questions

Exercise:

Problem:

What three factors will determine whether a nation has a higher or lower share of trade relative to its GDP?

Exercise:

Problem:

What is the difference between trade deficits and balance of trade?

Critical Thinking Questions

Exercise:

Problem:

Will nations that are more involved in foreign trade tend to have higher trade imbalances, lower trade imbalances, or is the pattern unpredictable?

Exercise:**Problem:**

Some economists warn that the persistent trade deficits and a negative current account balance that the United States has run will be a problem in the long run. Do you agree or not? Explain your answer.

Introduction to Exchange Rates and International Capital Flows

class="introduction"

Trade Around the World

Is a trade deficit
between the United
States and the
European Union
good or bad for the
U.S. economy?
(Credit: modification
of work by Milad
Mosapoor/Wikimedi
a Commons)



Note:

Is a Stronger Dollar Good for the U.S. Economy?

From 2002 to 2008, the U.S. dollar lost more than a quarter of its value in foreign currency markets. On January 1, 2002, one dollar was worth 1.11 euros. On April 24, 2008 it hit its lowest point with a dollar being worth 0.64 euros. During this period, the trade deficit between the United States and the European Union grew from a yearly total of approximately –85.7 billion dollars in 2002 to 95.8 billion dollars in 2008. Was this a good thing or a bad thing for the U.S. economy?

We live in a global world. U.S. consumers buy trillions of dollars worth of imported goods and services each year, not just from the European Union, but from all over the world. U.S. businesses sell trillions of dollars' worth of exports. U.S. citizens, businesses, and governments invest trillions of dollars abroad every year. Foreign investors, businesses, and governments invest trillions of dollars in the United States each year. Indeed, foreigners are a major buyer of U.S. federal debt.

Many people feel that a weaker dollar is bad for America, that it's an indication of a weak economy. But is it? This chapter will help answer that question.

Note:

Introduction to Exchange Rates and International Capital Flows

In this chapter, you will learn about:

- How the Foreign Exchange Market Works
- Demand and Supply Shifts in Foreign Exchange Markets
- Macroeconomic Effects of Exchange Rates
- Exchange Rate Policies

The world has over 150 different currencies, from the Afghanistan afghani and the Albanian lek all the way through the alphabet to the Zambian kwacha and the Zimbabwean dollar. For international economic transactions, households or firms will wish to exchange one currency for another. Perhaps the need for exchanging currencies will come from a German firm that exports products to Russia, but then wishes to exchange

the Russian rubles it has earned for euros, so that the firm can pay its workers and suppliers in Germany. Perhaps it will be a South African firm that wishes to purchase a mining operation in Angola, but to make the purchase it must convert South African rand to Angolan kwanza. Perhaps it will be an American tourist visiting China, who wishes to convert U.S. dollars to Chinese yuan to pay the hotel bill.

Exchange rates can sometimes change very swiftly. For example, in the United Kingdom the pound was worth \$2 in U.S. currency in spring 2008, but was worth only \$1.40 in U.S. currency six months later. For firms engaged in international buying, selling, lending, and borrowing, these swings in exchange rates can have an enormous effect on profits.

This chapter discusses the international dimension of money, which involves conversions from one currency to another at an exchange rate. An exchange rate is nothing more than a price—that is, the price of one currency in terms of another currency—and so they can be analyzed with the tools of supply and demand. The first module of this chapter begins with an overview of foreign exchange markets: their size, their main participants, and the vocabulary for discussing movements of exchange rates. The following module uses demand and supply graphs to analyze some of the main factors that cause shifts in exchange rates. A final module then brings the central bank and monetary policy back into the picture. Each country must decide whether to allow its exchange rate to be determined in the market, or have the central bank intervene in the exchange rate market. All the choices for exchange rate policy involve distinctive tradeoffs and risks.

How the Foreign Exchange Market Works

By the end of this section, you will be able to:

- Define "foreign exchange market"
- Describe different types of investments like foreign direct investments (FDI), portfolio investments, and hedging
- Explain how the appreciating or depreciating of currency affects exchange rates
- Identify who benefits from a stronger currency and benefits from a weaker currency

Most countries have different currencies, but not all. Sometimes small economies use the currency of an economically larger neighbor. For example, Ecuador, El Salvador, and Panama have decided to **dollarize**—that is, to use the U.S. dollar as their currency. Sometimes nations share a common currency. A large-scale example of a common currency is the decision by 17 European nations—including some very large economies such as France, Germany, and Italy—to replace their former currencies with the euro. With these exceptions duly noted, most of the international economy takes place in a situation of multiple national currencies in which both people and firms need to convert from one currency to another when selling, buying, hiring, borrowing, traveling, or investing across national borders. The market in which people or firms use one currency to purchase another currency is called the **foreign exchange market**.

You have encountered the basic concept of exchange rates in earlier chapters. In [The International Trade and Capital Flows](#), for example, we discussed how exchange rates are used to compare GDP statistics from countries where GDP is measured in different currencies. These earlier examples, however, took the actual exchange rate as given, as if it were a fact of nature. In reality, the exchange rate is a price—the price of one currency expressed in terms of units of another currency. The key framework for analyzing prices, whether in this course, any other economics course, in public policy, or business examples, is the operation of supply and demand in markets.

Note: Visit this [website](#) for an exchange rate calculator.

The Extraordinary Size of the Foreign Exchange Markets

The quantities traded in foreign exchange markets are breathtaking. A survey done in April, 2013 by the Bank of International Settlements, an international organization for banks and the financial industry, found that \$5.3 trillion *per day* was traded on foreign exchange markets, which makes the foreign exchange market the largest market in the world economy. In contrast, 2013 U.S. real GDP was \$15.8 trillion *per year*.

[\[link\]](#) shows the currencies most commonly traded on foreign exchange markets. The foreign exchange market is dominated by the U.S. dollar, the currencies used by nations in Western Europe (the euro, the British pound, and the Australian dollar), and the Japanese yen.

Currency	% Daily Share
U.S. dollar	87.0%
Euro	33.4%
Japanese yen	23.0%
British pound	11.8%
Australian dollar	8.6%
Swiss franc	5.2%

Currency	% Daily Share
Canadian dollar	4.6%
Mexican peso	2.5%
Chinese yuan	2.2%

Currencies Traded Most on Foreign Exchange Markets as of April, 2013(Source: <http://www.bis.org/publ/rpfx13fx.pdf>)

Demanders and Suppliers of Currency in Foreign Exchange Markets

In foreign exchange markets, demand and supply become closely interrelated, because a person or firm who demands one currency must at the same time supply another currency—and vice versa. To get a sense of this, it is useful to consider four groups of people or firms who participate in the market: (1) firms that are involved in international trade of goods and services; (2) tourists visiting other countries; (3) international investors buying ownership (or part-ownership) of a foreign firm; (4) international investors making financial investments that do not involve ownership. Let's consider these categories in turn.

Firms that buy and sell on international markets find that their costs for workers, suppliers, and investors are measured in the currency of the nation where their production occurs, but their revenues from sales are measured in the currency of the different nation where their sales happened. So, a Chinese firm exporting abroad will earn some other currency—say, U.S. dollars—but will need Chinese yuan to pay the workers, suppliers, and investors who are based in China. In the foreign exchange markets, this firm will be a supplier of U.S. dollars and a demander of Chinese yuan.

International tourists will supply their home currency to receive the currency of the country they are visiting. For example, an American tourist

who is visiting China will supply U.S. dollars into the foreign exchange market and demand Chinese yuan.

Financial investments that cross international boundaries, and require exchanging currency, are often divided into two categories. **Foreign direct investment (FDI)** refers to purchasing a firm (at least ten percent) in another country or starting up a new enterprise in a foreign country. For example, in 2008 the Belgian beer-brewing company InBev bought the U.S. beer-maker Anheuser-Busch for \$52 billion. To make this purchase of a U.S. firm, InBev would have to supply euros (the currency of Belgium) to the foreign exchange market and demand U.S. dollars.

The other kind of international financial investment, **portfolio investment**, involves a purely financial investment that does not entail any management responsibility. An example would be a U.S. financial investor who purchased bonds issued by the government of the United Kingdom, or deposited money in a British bank. To make such investments, the American investor would supply U.S. dollars in the foreign exchange market and demand British pounds.

Both foreign direct investment and portfolio investment involve an investor who supplies domestic currency and demands a foreign currency. With portfolio investment less than ten percent of a company is purchased. As such, portfolio investment is often made with a short term focus. With foreign direct investment more than ten percent of a company is purchased and the investor typically assumes some managerial responsibility; thus foreign direct investment tends to have a more long-run focus. As a practical matter, portfolio investments can be withdrawn from a country much more quickly than foreign direct investments. A U.S. portfolio investor who wants to buy or sell bonds issued by the government of the United Kingdom can do so with a phone call or a few clicks of a computer key. However, a U.S. firm that wants to buy or sell a company, such as one that manufactures automobile parts in the United Kingdom, will find that planning and carrying out the transaction takes a few weeks, even months. [\[link\]](#) summarizes the main categories of demanders and suppliers of currency.

Demand for the U.S. Dollar Comes from...	Supply of the U.S. Dollar Comes from...
A U.S. exporting firm that earned foreign currency and is trying to pay U.S.-based expenses	A foreign firm that has sold imported goods in the United States, earned U.S. dollars, and is trying to pay expenses incurred in its home country
Foreign tourists visiting the United States	U.S. tourists leaving to visit other countries
Foreign investors who wish to make direct investments in the U.S. economy	U.S. investors who want to make foreign direct investments in other countries
Foreign investors who wish to make portfolio investments in the U.S. economy	U.S. investors who want to make portfolio investments in other countries

The Demand and Supply Line-ups in Foreign Exchange Markets

Participants in the Exchange Rate Market

The foreign exchange market does not involve the ultimate suppliers and demanders of foreign exchange literally seeking each other out. If Martina decides to leave her home in Venezuela and take a trip in the United States, she does not need to find a U.S. citizen who is planning to take a vacation in Venezuela and arrange a person-to-person currency trade. Instead, the foreign exchange market works through financial institutions, and it operates on several levels.

Most people and firms who are exchanging a substantial quantity of currency go to a bank, and most banks provide foreign exchange as a

service to customers. These banks (and a few other firms), known as dealers, then trade the foreign exchange. This is called the interbank market.

In the world economy, roughly 2,000 firms are foreign exchange dealers. The U.S. economy has less than 100 foreign exchange dealers, but the largest 12 or so dealers carry out more than half the total transactions. The foreign exchange market has no central location, but the major dealers keep a close watch on each other at all times.

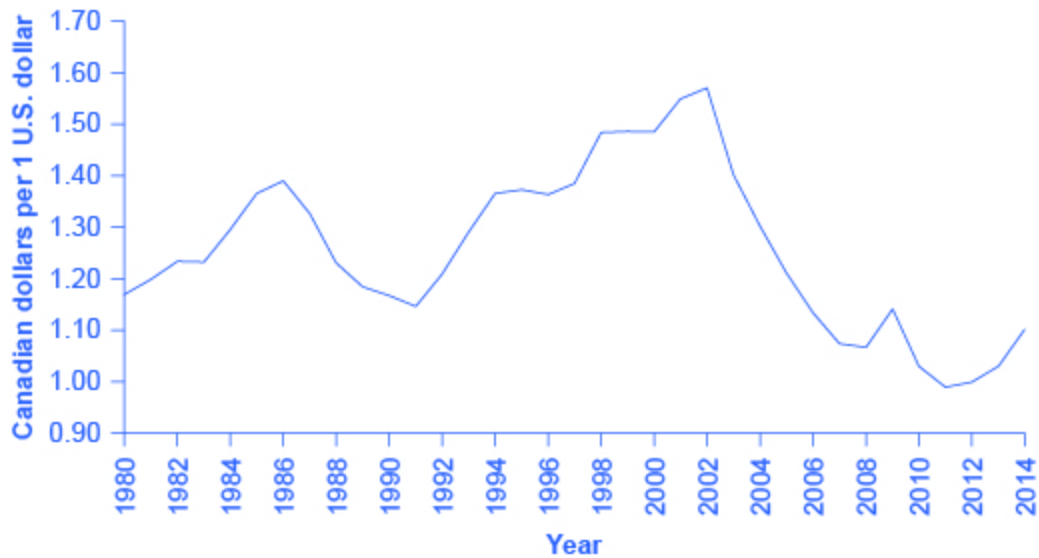
The foreign exchange market is huge not because of the demands of tourists, firms, or even foreign direct investment, but instead because of portfolio investment and the actions of interlocking foreign exchange dealers. International tourism is a very large industry, involving about \$1 trillion per year. Global exports are about 23% of global GDP; which is about \$18 trillion per year. Foreign direct investment totaled about \$1.5 trillion in the end of 2013. These quantities are dwarfed, however, by the \$5.3 trillion *per day* being traded in foreign exchange markets. Most transactions in the foreign exchange market are for portfolio investment—relatively short-term movements of financial capital between currencies—and because of the actions of the large foreign exchange dealers as they constantly buy and sell with each other.

Strengthening and Weakening Currency

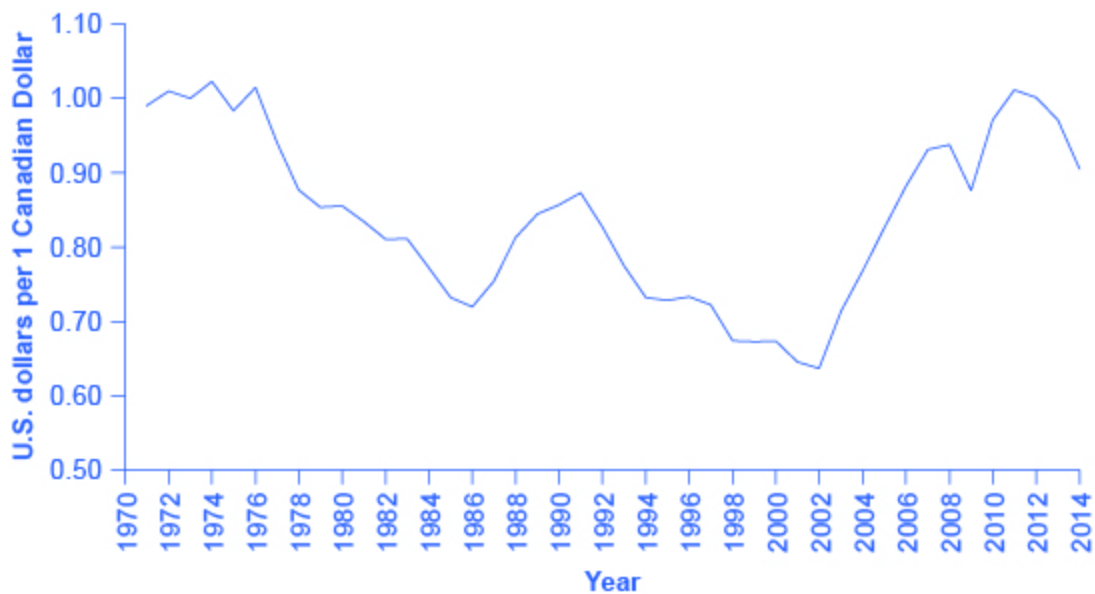
When the prices of most goods and services change, the price is said to “rise” or “fall.” For exchange rates, the terminology is different. When the exchange rate for a currency rises, so that the currency exchanges for more of other currencies, it is referred to as **appreciating** or “strengthening.” When the exchange rate for a currency falls, so that a currency trades for less of other currencies, it is referred to as **depreciating** or “weakening.”

To illustrate the use of these terms, consider the exchange rate between the U.S. dollar and the Canadian dollar since 1980, shown in [\[link\]](#) (a). The vertical axis in [\[link\]](#) (a) shows the price of \$1 in U.S. currency, measured in terms of Canadian currency. Clearly, exchange rates can move up and down substantially. A U.S. dollar traded for \$1.17 Canadian in 1980. The

U.S. dollar appreciated or strengthened to \$1.39 Canadian in 1986, depreciated or weakened to \$1.15 Canadian in 1991, and then appreciated or strengthened to \$1.60 Canadian by early in 2002, fell to roughly \$1.20 Canadian in 2009, and then had a sharp spike up and decline in 2009 and 2010. The units in which exchange rates are measured can be confusing, because the exchange rate of the U.S. dollar is being measured using a different currency—the Canadian dollar. But exchange rates always measure the price of one unit of currency by using a different currency. Strengthen or Appreciate vs. Weaken or Depreciate



(a) U.S. dollar exchange rate in Canadian dollars



(b) Canadian dollar exchange rate in U.S. dollars

Exchange rates tend to fluctuate substantially, even between bordering countries such as the United States and Canada. By looking closely at the time values (the years vary slightly on these graphs), it is clear that the values in part (a) are a mirror image of part (b), which demonstrates that the depreciation of one currency correlates to the appreciation of the other and vice versa. This means that when comparing the exchange rates between two

countries (in this case, the United States and Canada), the depreciation (or weakening) of one country (the U.S. dollar for this example) indicates the appreciation (or strengthening) of the other currency (which in this example is the Canadian dollar). (Source: Federal Reserve Economic Data (FRED) (a) <https://research.stlouisfed.org/fred2/series/EXCAUS> ; (b) <https://research.stlouisfed.org/fred2/series/CCUSSP01CAQ650N>)

In looking at the exchange rate between two currencies, the appreciation or strengthening of one currency must mean the depreciation or weakening of the other. [\[link\]](#) (b) shows the exchange rate for the Canadian dollar, measured in terms of U.S. dollars. The exchange rate of the U.S. dollar measured in Canadian dollars, shown in [\[link\]](#) (a), is a perfect mirror image with the exchange rate of the Canadian dollar measured in U.S. dollars, shown in [\[link\]](#) (b). A fall in the Canada \$/U.S. \$ ratio means a rise in the U.S. \$/Canada \$ ratio, and vice versa.

With the price of a typical good or service, it is clear that higher prices benefit sellers and hurt buyers, while lower prices benefit buyers and hurt sellers. In the case of exchange rates, where the buyers and sellers are not always intuitively obvious, it is useful to trace through how different participants in the market will be affected by a stronger or weaker currency. Consider, for example, the impact of a stronger U.S. dollar on six different groups of economic actors, as shown in [\[link\]](#): (1) U.S. exporters selling abroad; (2) foreign exporters (that is, firms selling imports in the U.S. economy); (3) U.S. tourists abroad; (4) foreign tourists visiting the United States; (5) U.S. investors (either foreign direct investment or portfolio investment) considering opportunities in other countries; (6) and foreign investors considering opportunities in the U.S. economy.

How Do Exchange Rate Movements Affect Each Group?

	A Stronger U.S. Dollar	A Weaker U.S. Dollar
A U.S. exporting firm		
A foreign firm exporting to the United States		
A U.S. tourist abroad		
A foreign tourist in the United States		
A U.S. investor abroad		
A foreign investor in the United States		

Exchange rate movements affect exporters, tourists, and international investors in different ways.

For a U.S. firm selling abroad, a stronger U.S. dollar is a curse. A strong U.S. dollar means that foreign currencies are correspondingly weak. When this exporting firm earns foreign currencies through its export sales, and then converts them back to U.S. dollars to pay workers, suppliers, and investors, the stronger dollar means that the foreign currency buys fewer U.S. dollars than if the currency had not strengthened, and that the firm's profits (as measured in dollars) fall. As a result, the firm may choose to reduce its exports, or it may raise its selling price, which will also tend to reduce its exports. In this way, a stronger currency reduces a country's exports.

Conversely, for a foreign firm selling in the U.S. economy, a stronger dollar is a blessing. Each dollar earned through export sales, when traded back into the home currency of the exporting firm, will now buy more of the home currency than expected before the dollar had strengthened. As a result, the stronger dollar means that the importing firm will earn higher

profits than expected. The firm will seek to expand its sales in the U.S. economy, or it may reduce prices, which will also lead to expanded sales. In this way, a stronger U.S. dollar means that consumers will purchase more from foreign producers, expanding the country's level of imports.

For a U.S. tourist abroad, who is exchanging U.S. dollars for foreign currency as necessary, a stronger U.S. dollar is a benefit. The tourist receives more foreign currency for each U.S. dollar, and consequently the cost of the trip in U.S. dollars is lower. When a country's currency is strong, it is a good time for citizens of that country to tour abroad. Imagine a U.S. tourist who has saved up \$5,000 for a trip to South Africa. In January 2008, \$1 bought 7 South African rand, so the tourist had 35,000 rand to spend. In January 2009, \$1 bought 10 rand, so the tourist had 50,000 rand to spend. By January 2010, \$1 bought only 7.5 rand. Clearly, 2009 was the year for U.S. tourists to visit South Africa. For foreign visitors to the United States, the opposite pattern holds true. A relatively stronger U.S. dollar means that their own currencies are relatively weaker, so that as they shift from their own currency to U.S. dollars, they have fewer U.S. dollars than previously. When a country's currency is strong, it is not an especially good time for foreign tourists to visit.

A stronger dollar injures the prospects of a U.S. financial investor who has already invested money in another country. A U.S. financial investor abroad must first convert U.S. dollars to a foreign currency, invest in a foreign country, and then later convert that foreign currency back to U.S. dollars. If in the meantime the U.S. dollar becomes stronger and the foreign currency becomes weaker, then when the investor converts back to U.S. dollars, the rate of return on that investment will be less than originally expected at the time it was made.

However, a stronger U.S. dollar boosts the returns of a foreign investor putting money into a U.S. investment. That foreign investor converts from the home currency to U.S. dollars and seeks a U.S. investment, while later planning to switch back to the home currency. If, in the meantime, the dollar grows stronger, then when the time comes to convert from U.S. dollars back to the foreign currency, the investor will receive more foreign currency than expected at the time the original investment was made.

The preceding paragraphs all focus on the case where the U.S. dollar becomes stronger. The corresponding happy or unhappy economic reactions are illustrated in the first column of [\[link\]](#).

At this point, you should have a good sense of the major players in the foreign exchange market: firms involved in international trade, tourists, international financial investors, banks, and foreign exchange dealers. The next module shows how the tools of demand and supply can be used in foreign exchange markets to explain the underlying causes of stronger and weaker currencies (“stronger” and “weaker” addressed more in the following Clear It Up feature).

Note:

Why is a stronger currency not necessarily better?

One common misunderstanding about exchange rates is that a “stronger” or “appreciating” currency must be better than a “weaker” or “depreciating” currency. After all, is it not obvious that “strong” is better than “weak”? But do not let the terminology confuse you. When a currency becomes stronger, so that it purchases more of other currencies, it benefits some in the economy and injures others. Stronger currency is not necessarily better, it is just different.

Key Concepts and Summary

In the foreign exchange market, people and firms exchange one currency to purchase another currency. The demand for dollars comes from those U.S. export firms seeking to convert their earnings in foreign currency back into U.S. dollars; foreign tourists converting their earnings in a foreign currency back into U.S. dollars; and foreign investors seeking to make financial investments in the U.S. economy. On the supply side of the foreign exchange market for the trading of U.S. dollars are foreign firms that have sold imports in the U.S. economy and are seeking to convert their earnings back to their home currency; U.S. tourists abroad; and U.S. investors seeking to make financial investments in foreign economies. When

currency A can buy more of currency B, then currency A has strengthened or appreciated relative to B. When currency A can buy less of currency B, then currency A has weakened or depreciated relative to B. If currency A strengthens or appreciates relative to currency B, then currency B must necessarily weaken or depreciate with regard to currency A. A stronger currency benefits those who are buying with that currency and injures those who are selling. A weaker currency injures those, like importers, who are buying with that currency and benefits those who are selling with it, like exporters.

Self-Check Questions

Exercise:

Problem:

How will a stronger euro affect the following economic agents?

- a. A British exporter to Germany.
- b. A Dutch tourist visiting Chile.
- c. A Greek bank investing in a Canadian government bond.
- d. A French exporter to Germany.

Solution:

- a. The British use the pound sterling, while Germans use the euro, so a British exporter will receive euros from export sales, which will need to be exchanged for pounds. A stronger euro will mean more pounds per euro, so the exporter will be better off. In addition, the lower price for German imports will stimulate demand for British exports. For both these reasons, a stronger euro benefits the British exporter.
- b. The Dutch use euros while the Chileans use pesos, so the Dutch tourist needs to turn euros into Chilean pesos. An increase in the euro means that the tourist will get more pesos per euro. As a consequence, the Dutch tourist will have a less expensive vacation than he planned, so the tourist will be better off.

- c. The Greek use euros while the Canadians use dollars. An increase in the euro means it will buy more Canadian dollars. As a result, the Greek bank will see a decrease in the cost of the Canadian bonds, so it may purchase more bonds. Either way, the Greek bank benefits.
- d. Since both the French and Germans use the euro, an increase in the euro, in terms of other currencies, should have no impact on the French exporter.

Review Questions

Exercise:

Problem: What is the foreign exchange market?

Exercise:

Problem:

Describe some buyers and some sellers in the market for U.S. dollars.

Exercise:

Problem:

What is the difference between foreign direct investment and portfolio investment?

Exercise:

Problem: What does it mean to hedge a financial transaction?

Exercise:

Problem:

What does it mean to say that a currency appreciates? Depreciates? Becomes stronger? Becomes weaker?

Critical Thinking Question

Exercise:

Problem:

Why would a nation “dollarize”—that is, adopt another country’s currency instead of having its own?

Exercise:

Problem:

Can you think of any major disadvantages to dollarization? How would a central bank work in a country that has dollarized?

Problems

Exercise:

Problem:

A British pound cost \$1.56 in U.S. dollars in 1996, but \$1.66 in U.S. dollars in 1998. Was the pound weaker or stronger against the dollar? Did the dollar appreciate or depreciate versus the pound?

Exercise:

Problem:

In [\[link\]](#) calculate the cost of a U.S. dollar in terms of British pounds in 1996 and 1998.

Glossary

appreciating

when a currency is worth more in terms of other currencies; also called “strengthening”

depreciating

when a currency is worth less in terms of other currencies; also called “weakening”

dollarize

a country that is not the United States uses the U.S. dollar as its currency

foreign direct investment (FDI)

purchasing more than ten percent of a firm or starting a new enterprise in another country

foreign exchange market

the market in which people use one currency to buy another currency

hedge

using a financial transaction as protection against risk

portfolio investment

an investment in another country that is purely financial and does not involve any management responsibility

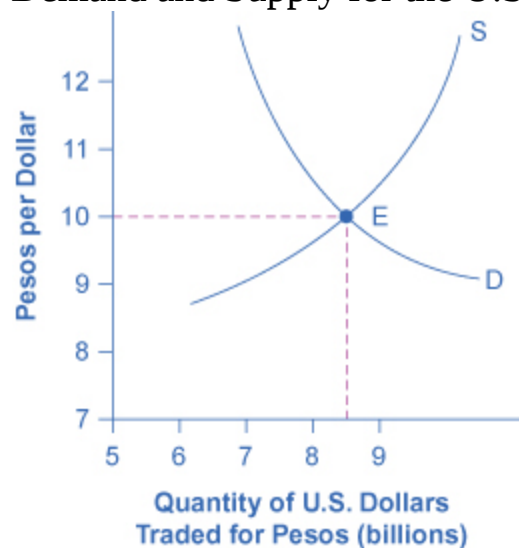
Demand and Supply Shifts in Foreign Exchange Markets

By the end of this section, you will be able to:

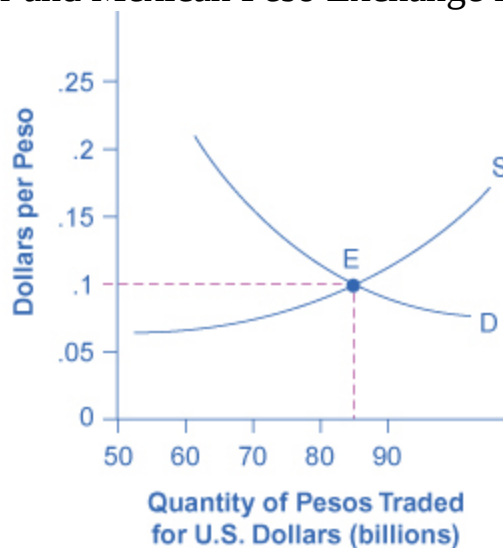
- Explain supply and demand for exchange rates
- Define arbitrage
- Explain purchasing power parity's importance when comparing countries.

The foreign exchange market involves firms, households, and investors who demand and supply currencies coming together through their banks and the key foreign exchange dealers. [\[link\]](#) (a) offers an example for the exchange rate between the U.S. dollar and the Mexican peso. The vertical axis shows the exchange rate for U.S. dollars, which in this case is measured in pesos. The horizontal axis shows the quantity of U.S. dollars being traded in the foreign exchange market each day. The demand curve (D) for U.S. dollars intersects with the supply curve (S) of U.S. dollars at the equilibrium point (E), which is an exchange rate of 10 pesos per dollar and a total volume of \$8.5 billion.

Demand and Supply for the U.S. Dollar and Mexican Peso Exchange Rate



(a) The U.S. dollar exchange rate



(b) The Mexican peso exchange rate

(a) The quantity measured on the horizontal axis is in U.S. dollars, and the exchange rate on the vertical axis is the price of U.S. dollars measured in Mexican pesos. (b) The quantity measured on the horizontal axis is in Mexican pesos, while the

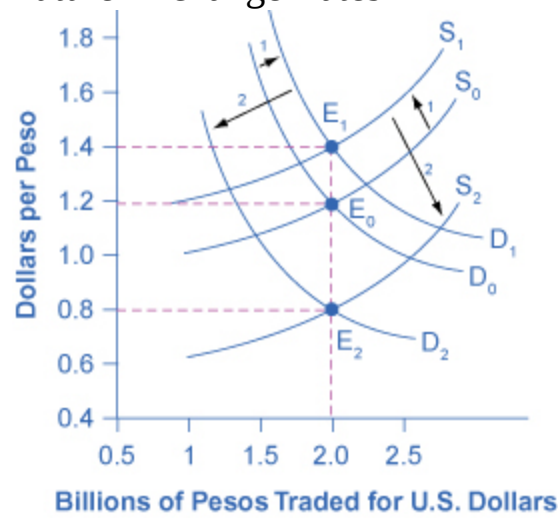
price on the vertical axis is the price of pesos measured in U.S. dollars. In both graphs, the equilibrium exchange rate occurs at point E, at the intersection of the demand curve (D) and the supply curve (S).

[\[link\]](#) (b) presents the same demand and supply information from the perspective of the Mexican peso. The vertical axis shows the exchange rate for Mexican pesos, which is measured in U.S. dollars. The horizontal axis shows the quantity of Mexican pesos traded in the foreign exchange market. The demand curve (D) for Mexican pesos intersects with the supply curve (S) of Mexican pesos at the equilibrium point (E), which is an exchange rate of 10 cents in U.S. currency for each Mexican peso and a total volume of 85 billion pesos. Note that the two exchange rates are inverses: 10 pesos per dollar is the same as 10 cents per peso (or \$0.10 per peso). In the actual foreign exchange market, almost all of the trading for Mexican pesos is done for U.S. dollars. What factors would cause the demand or supply to shift, thus leading to a change in the equilibrium exchange rate? The answer to this question is discussed in the following section.

Expectations about Future Exchange Rates

One reason to demand a currency on the foreign exchange market is the belief that the value of the currency is about to increase. One reason to supply a currency—that is, sell it on the foreign exchange market—is the expectation that the value of the currency is about to decline. For example, imagine that a leading business newspaper, like the *Wall Street Journal* or the *Financial Times*, runs an article predicting that the Mexican peso will appreciate in value. The likely effects of such an article are illustrated in [\[link\]](#). Demand for the Mexican peso shifts to the right, from D_0 to D_1 , as investors become eager to purchase pesos. Conversely, the supply of pesos shifts to the left, from S_0 to S_1 , because investors will be less willing to give them up. The result is that the equilibrium exchange rate rises from 10 cents/peso to 12 cents/peso and the equilibrium exchange rate rises from 85 billion to 90 billion pesos as the equilibrium moves from E_0 to E_1 .

Exchange Rate Market for Mexican Peso Reacts to Expectations about Future Exchange Rates



An announcement that the peso exchange rate is likely to strengthen in the future will lead to greater demand for the peso in the present from investors who wish to benefit from the appreciation. Similarly, it will make investors less likely to supply pesos to the foreign exchange market. Both the shift of demand to the right and the shift of supply to the left cause an immediate appreciation in the exchange rate.

[\[link\]](#) also illustrates some peculiar traits of supply and demand diagrams in the foreign exchange market. In contrast to all the other cases of supply and demand you have considered, in the foreign exchange market, supply and demand typically both move at the same time. Groups of participants in the

foreign exchange market like firms and investors include some who are buyers and some who are sellers. An expectation of a future shift in the exchange rate affects both buyers and sellers—that is, it affects both demand and supply for a currency.

The shifts in demand and supply curves both cause the exchange rate to shift in the same direction; in this example, they both make the peso exchange rate stronger. However, the shifts in demand and supply work in opposing directions on the quantity traded. In this example, the rising demand for pesos is causing the quantity to rise while the falling supply of pesos is causing quantity to fall. In this specific example, the result is a higher quantity. But in other cases, the result could be that quantity remains unchanged or declines.

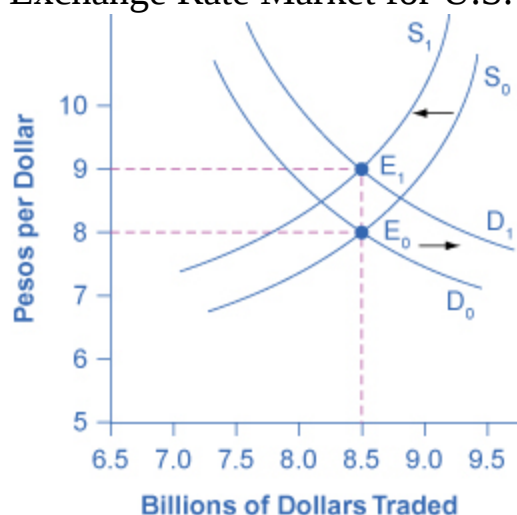
This example also helps to explain why exchange rates often move quite substantially in a short period of a few weeks or months. When investors expect a country's currency to strengthen in the future, they buy the currency and cause it to appreciate immediately. The appreciation of the currency can lead other investors to believe that future appreciation is likely—and thus lead to even further appreciation. Similarly, a fear that a currency *might* weaken quickly leads to an *actual* weakening of the currency, which often reinforces the belief that the currency is going to weaken further. Thus, beliefs about the future path of exchange rates can be self-reinforcing, at least for a time, and a large share of the trading in foreign exchange markets involves dealers trying to outguess each other on what direction exchange rates will move next.

Differences across Countries in Rates of Return

The motivation for investment, whether domestic or foreign, is to earn a return. If rates of return in a country look relatively high, then that country will tend to attract funds from abroad. Conversely, if rates of return in a country look relatively low, then funds will tend to flee to other economies. Changes in the expected rate of return will shift demand and supply for a currency. For example, imagine that interest rates rise in the United States as compared with Mexico. Thus, financial investments in the United States promise a higher return than they previously did. As a result, more investors

will demand U.S. dollars so that they can buy interest-bearing assets and fewer investors will be willing to supply U.S. dollars to foreign exchange markets. Demand for the U.S. dollar will shift to the right, from D_0 to D_1 , and supply will shift to the left, from S_0 to S_1 , as shown in [\[link\]](#). The new equilibrium (E_1), will occur at an exchange rate of nine pesos/dollar and the same quantity of \$8.5 billion. Thus, a higher interest rate or rate of return relative to other countries leads a nation's currency to appreciate or strengthen, and a lower interest rate relative to other countries leads a nation's currency to depreciate or weaken. Since a nation's central bank can use monetary policy to affect its interest rates, a central bank can also cause changes in exchange rates—a connection that will be discussed in more detail later in this chapter.

Exchange Rate Market for U.S. Dollars Reacts to Higher Interest Rates



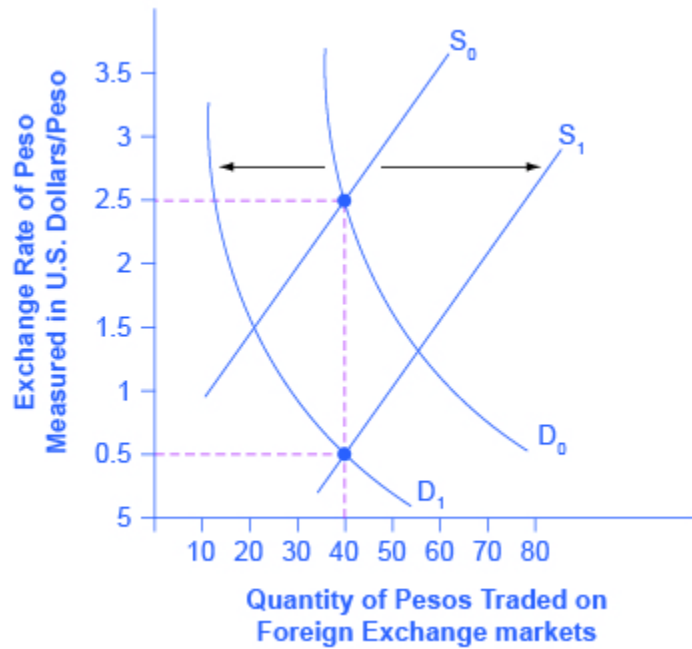
A higher rate of return for U.S. dollars makes holding dollars more attractive. Thus, the demand for dollars in the foreign exchange market shifts to the right, from D_0 to D_1 , while the supply of dollars shifts to the left, from S_0 to S_1 . The new equilibrium (E_1) has a stronger exchange rate

than the original equilibrium (E_0), but in this example, the equilibrium quantity traded does not change.

Relative Inflation

If a country experiences a relatively high inflation rate compared with other economies, then the buying power of its currency is eroding, which will tend to discourage anyone from wanting to acquire or to hold the currency. [\[link\]](#) shows an example based on an actual episode concerning the Mexican peso. In 1986–87, Mexico experienced an inflation rate of over 200%. Not surprisingly, as inflation dramatically decreased the purchasing power of the peso in Mexico, the exchange rate value of the peso declined as well. As shown in [\[link\]](#), demand for the peso on foreign exchange markets decreased from D_0 to D_1 , while supply of the peso increased from S_0 to S_1 . The equilibrium exchange rate fell from \$2.50 per peso at the original equilibrium (E_0) to \$0.50 per peso at the new equilibrium (E_1). In this example, the quantity of pesos traded on foreign exchange markets remained the same, even as the exchange rate shifted.

Exchange Rate Markets React to Higher Inflation



If a currency is experiencing relatively high inflation, then its buying power is decreasing and international investors will be less eager to hold it. Thus, a rise in inflation in the Mexican peso would lead demand to shift from D_0 to D_1 , and supply to increase from S_0 to S_1 . Both movements in demand and supply would cause the currency to depreciate. The effect on the quantity traded is drawn here as a decrease, but in truth it could be an increase or no change, depending on the actual movements of demand and supply.

Note: Visit this [website](#) to learn about the Big Mac index.

Purchasing Power Parity

Over the long term, exchange rates must bear some relationship to the buying power of the currency in terms of goods that are internationally traded. If at a certain exchange rate it was much cheaper to buy internationally traded goods—such as oil, steel, computers, and cars—in one country than in another country, businesses would start buying in the cheap country, selling in other countries, and pocketing the profits.

For example, if a U.S. dollar is worth \$1.60 in Canadian currency, then a car that sells for \$20,000 in the United States should sell for \$32,000 in Canada. If the price of cars in Canada was much lower than \$32,000, then at least some U.S. car-buyers would convert their U.S. dollars to Canadian dollars and buy their cars in Canada. If the price of cars was much higher than \$32,000 in this example, then at least some Canadian buyers would convert their Canadian dollars to U.S. dollars and go to the United States to purchase their cars. This is known as **arbitrage**, the process of buying and selling goods or currencies across international borders at a profit. It may occur slowly, but over time, it will force prices and exchange rates to align so that the price of internationally traded goods is similar in all countries.

The exchange rate that equalizes the prices of internationally traded goods across countries is called the **purchasing power parity (PPP)** exchange rate. A group of economists at the International Comparison Program, run by the World Bank, have calculated the PPP exchange rate for all countries, based on detailed studies of the prices and quantities of internationally tradable goods.

The purchasing power parity exchange rate has two functions. First, PPP exchange rates are often used for international comparison of GDP and other economic statistics. Imagine that you are preparing a table showing the size of GDP in many countries in several recent years, and for ease of comparison, you are converting all the values into U.S. dollars. When you insert the value for Japan, you need to use a yen/dollar exchange rate. But should you use the market exchange rate or the PPP exchange rate? Market exchange rates bounce around. In summer 2008, the exchange rate was 108 yen/dollar, but in late 2009 the U.S. dollar exchange rate versus the yen was

90 yen/dollar. For simplicity, say that Japan's GDP was ¥500 trillion in both 2008 and 2009. If you use the market exchange rates, then Japan's GDP will be \$4.6 trillion in 2008 (that is, ¥500 trillion / (¥108/dollar)) and \$5.5 trillion in 2009 (that is, ¥500 trillion / (¥90/dollar)).

Of course, it is not true that Japan's economy increased enormously in 2009—in fact, Japan had a recession like much of the rest of the world. The misleading appearance of a booming Japanese economy occurs only because we used the market exchange rate, which often has short-run rises and falls. However, PPP exchange rates stay fairly constant and change only modestly, if at all, from year to year.

The second function of PPP is that exchange rates will often get closer and closer to it as time passes. It is true that in the short run and medium run, as exchange rates adjust to relative inflation rates, rates of return, and to expectations about how interest rates and inflation will shift, the exchange rates will often move away from the PPP exchange rate for a time. But, knowing the PPP will allow you to track and predict exchange rate relationships.

Key Concepts and Summary

In the extreme short run, ranging from a few minutes to a few weeks, exchange rates are influenced by speculators who are trying to invest in currencies that will grow stronger, and to sell currencies that will grow weaker. Such speculation can create a self-fulfilling prophecy, at least for a time, where an expected appreciation leads to a stronger currency and vice versa. In the relatively short run, exchange rate markets are influenced by differences in rates of return. Countries with relatively high real rates of return (for example, high interest rates) will tend to experience stronger currencies as they attract money from abroad, while countries with relatively low rates of return will tend to experience weaker exchange rates as investors convert to other currencies.

In the medium run of a few months or a few years, exchange rate markets are influenced by inflation rates. Countries with relatively high inflation will tend to experience less demand for their currency than countries with

lower inflation, and thus currency depreciation. Over long periods of many years, exchange rates tend to adjust toward the purchasing power parity (PPP) rate, which is the exchange rate such that the prices of internationally tradable goods in different countries, when converted at the PPP exchange rate to a common currency, are similar in all economies.

Self-Check Questions

Exercise:

Problem:

Suppose that political unrest in Egypt leads financial markets to anticipate a depreciation in the Egyptian pound. How will that affect the demand for pounds, supply of pounds, and exchange rate for pounds compared to, say, U.S. dollars?

Solution:

Expected depreciation in a currency will lead people to divest themselves of the currency. We should expect to see an increase in the supply of pounds and a decrease in demand for pounds. The result should be a decrease in the value of the pound *vis à vis* the dollar.

Exercise:

Problem:

Suppose U.S. interest rates decline compared to the rest of the world. What would be the likely impact on the demand for dollars, supply of dollars, and exchange rate for dollars compared to, say, euros?

Solution:

Lower U.S. interest rates make U.S. assets less desirable compared to assets in the European Union. We should expect to see a decrease in demand for dollars and an increase in supply of dollars in foreign currency markets. As a result, we should expect to see the dollar depreciate compared to the euro.

Exercise:**Problem:**

Suppose Argentina gets inflation under control and the Argentine inflation rate decreases substantially. What would likely happen to the demand for Argentine pesos, the supply of Argentine pesos, and the peso/U.S. dollar exchange rate?

Solution:

A decrease in Argentine inflation relative to other countries should cause an increase in demand for pesos, a decrease in supply of pesos, and an appreciation of the peso in foreign currency markets.

Review Questions**Exercise:****Problem:**

Does an expectation of a stronger exchange rate in the future affect the exchange rate in the present? If so, how?

Exercise:**Problem:**

Does a higher rate of return in a nation's economy, all other things being equal, affect the exchange rate of its currency? If so, how?

Exercise:**Problem:**

Does a higher inflation rate in an economy, other things being equal, affect the exchange rate of its currency? If so, how?

Exercise:

Problem: What is the purchasing power parity exchange rate?

Critical Thinking Questions

Exercise:

Problem:

If a country's currency is expected to appreciate in value, what would you think will be the impact of expected exchange rates on yields (e.g., the interest rate paid on government bonds) in that country? *Hint:* Think about how expected exchange rate changes and interest rates affect demand and supply for a currency.

Exercise:

Problem:

Do you think that a country experiencing hyperinflation is more or less likely to have an exchange rate equal to its purchasing power parity value when compared to a country with a low inflation rate?

Glossary

arbitrage

the process of buying a good and selling goods across borders to take advantage of international price differences

purchasing power parity (PPP)

the exchange rate that equalizes the prices of internationally traded goods across countries

Macroeconomic Effects of Exchange Rates

By the end of this section you will be able to:

- Explain how exchange rate shifting influences aggregate demand and supply
- Explain how loans and banks can also be influenced by shifting exchange rates

A central bank will be concerned about the exchange rate for multiple reasons: (1) Movements in the exchange rate will affect the quantity of aggregate demand in an economy; (2) frequent substantial fluctuations in the exchange rate can disrupt international trade and cause problems in a nation's banking system—this may contribute to an unsustainable balance of trade and large inflows of international financial capital, which can set the economy up for a deep recession if international investors decide to move their money to another country. Let's discuss these scenarios in turn.

Exchange Rates, Aggregate Demand, and Aggregate Supply

Foreign trade in goods and services typically involves incurring the costs of production in one currency while receiving revenues from sales in another currency. As a result, movements in exchange rates can have a powerful effect on incentives to export and import, and thus on aggregate demand in the economy as a whole.

For example, in 1999, when the euro first became a currency, its value measured in U.S. currency was \$1.06/euro. By the end of 2013, the euro had risen (and the U.S. dollar had correspondingly weakened) to \$1.37/euro. Consider the situation of a French firm that each year incurs €10 million in costs, and sells its products in the United States for \$10 million. In 1999, when this firm converted \$10 million back to euros at the exchange rate of \$1.06/euro (that is, $\$10 \text{ million} \times [\text{€}1/\$1.06]$), it received €9.4 million, and suffered a loss. In 2013, when this same firm converted \$10 million back to euros at the exchange rate of \$1.37/euro (that is, $\$10 \text{ million} \times [\text{€}1 \text{ euro}/\$1.37]$), it received approximately €7.3 million and an even larger loss. This example shows how a stronger euro discourages

exports by the French firm, because it makes the costs of production in the domestic currency higher relative to the sales revenues earned in another country. From the point of view of the U.S. economy, the example also shows how a weaker U.S. dollar encourages exports.

Since an increase in exports results in more dollars flowing into the economy, and an increase in imports means more dollars are flowing out, it is easy to conclude that exports are “good” for the economy and imports are “bad,” but this overlooks the role of exchange rates. If an American consumer buys a Japanese car for \$20,000 instead of an American car for \$30,000, it may be tempting to argue that the American economy has lost out. However, the Japanese company will have to convert those dollars to yen to pay its workers and operate its factories. Whoever buys those dollars will have to use them to purchase American goods and services, so the money comes right back into the American economy. At the same time, the consumer saves money by buying a less expensive import, and can use the extra money for other purposes.

Fluctuations in Exchange Rates

Exchange rates can fluctuate a great deal in the short run. As yet one more example, the Indian rupee moved from 39 rupees/dollar in February 2008 to 51 rupees/dollar in March 2009, a decline of more than one-fourth in the value of the rupee on foreign exchange markets. [\[link\]](#) earlier showed that even two economically developed neighboring economies like the United States and Canada can see significant movements in exchange rates over a few years. For firms that depend on export sales, or firms that rely on imported inputs to production, or even purely domestic firms that compete with firms tied into international trade—which in many countries adds up to half or more of a nation’s GDP—sharp movements in exchange rates can lead to dramatic changes in profits and losses. So, a central bank may desire to keep exchange rates from moving too much as part of providing a stable business climate, where firms can focus on productivity and innovation, not on reacting to exchange rate fluctuations.

Banks play a vital role in any economy in facilitating transactions and in making loans to firms and consumers. When most of a country’s largest

banks become bankrupt simultaneously, a sharp decline in aggregate demand and a deep recession results. Since the main responsibilities of a central bank are to control the money supply and to ensure that the banking system is stable, a central bank must be concerned about whether large and unexpected exchange rate depreciation will drive most of the country's existing banks into bankruptcy. For more on this concern, return to the chapter on [The International Trade and Capital Flows](#).

Summing Up Public Policy and Exchange Rates

Every nation would prefer a stable exchange rate to facilitate international trade and reduce the degree of risk and uncertainty in the economy. However, a nation may sometimes want a weaker exchange rate to stimulate aggregate demand and reduce a recession, or a stronger exchange rate to fight inflation. The country must also be concerned that rapid movements from a weak to a strong exchange rate may cripple its export industries, while rapid movements from a strong to a weak exchange rate can cripple its banking sector. In short, every choice of an exchange rate—whether it should be stronger or weaker, or fixed or changing—represents potential tradeoffs.

Key Concepts and Summary

A central bank will be concerned about the exchange rate for several reasons. Exchange rates will affect imports and exports, and thus affect aggregate demand in the economy. Fluctuations in exchange rates may cause difficulties for many firms, but especially banks. The exchange rate may accompany unsustainable flows of international financial capital.

Self-Check Questions

Exercise:

Problem:

This chapter has explained that “one of the most economically destructive effects of exchange rate fluctuations can happen through the banking system,” if banks borrow from abroad to lend domestically. Why is this less likely to be a problem for the U.S. banking system?

Solution:

The problem occurs when banks borrow foreign currency but lend in domestic currency. Since banks' assets (loans they made) are in domestic currency, while their debts (money they borrowed) are in foreign currency, when the domestic currency declines, their debts grow larger. If the domestic currency falls substantially in value, as happened during the Asian financial crisis, then the banking system could fail. This problem is unlikely to occur for U.S. banks because, even when they borrow from abroad, they tend to borrow dollars. Remember, there are trillions of dollars in circulation in the global economy. Since both assets and debts are in dollars, a change in the value of the dollar does not cause banking system failure the way it can when banks borrow in foreign currency.

Exercise:**Problem:**

A booming economy can attract financial capital inflows, which promote further growth. But capital can just as easily flow out of the country, leading to economic recession. Is a country whose economy is booming because it decided to stimulate consumer spending more or less likely to experience capital flight than an economy whose boom is caused by economic investment expenditure?

Solution:

While capital flight is possible in either case, if a country borrows to invest in real capital it is more likely to be able to generate the income

to pay back its debts than a country that borrows to finance consumption. As a result, an investment-stimulated economy is less likely to provoke capital flight and economic recession.

Review Questions

Exercise:

Problem:

What are some of the reasons a central bank is likely to care, at least to some extent, about the exchange rate?

Exercise:

Problem:

How can an unexpected fall in exchange rates injure the financial health of a nation's banks?

Critical Thinking Questions

Exercise:

Problem:

Suppose a country has an overall balance of trade so that exports of goods and services equal imports of goods and services. Does that imply that the country has balanced trade with *each* of its trading partners?

Exercise:

Problem:

We learned that monetary policy is amplified by changes in exchange rates and the corresponding changes in the balance of trade. From the perspective of a nation's central bank, is this a good thing or a bad thing?

Exercise:**Problem:**

If a developing country needs foreign capital inflows, management expertise, and technology, how can it encourage foreign investors while at the same time protect itself against capital flight and banking system collapse, as happened during the Asian financial crisis?

Exchange Rate Policies

By the end of this section, you will be able to:

- Differentiate among a floating exchange rate, a soft peg, a hard peg, and a merged currency
- Identify the tradeoffs that come with a floating exchange rate, a soft peg, a hard peg, and a merged currency

Exchange rate policies come in a range of different forms listed in [\[link\]](#): let the foreign exchange market determine the exchange rate; let the market set the value of the exchange rate most of the time, but have the central bank sometimes intervene to prevent fluctuations that seem too large; have the central bank guarantee a specific exchange rate; or share a currency with other countries. Let's discuss each type of exchange rate policy and its tradeoffs.

A Spectrum of Exchange Rate Policies



A nation may adopt one of a variety of exchange rate regimes, from floating rates in which the foreign exchange market determines the rates to pegged rates where governments intervene to manage the value of the exchange rate, to a common currency where the nation adopts the currency of another country or group of countries.

Floating Exchange Rates

A policy which allows the foreign exchange market to set exchange rates is referred to as a **floating exchange rate**. The U.S. dollar is a floating

exchange rate, as are the currencies of about 40% of the countries in the world economy. The major concern with this policy is that exchange rates can move a great deal in a short time.

Consider the U.S. exchange rate expressed in terms of another fairly stable currency, the Japanese yen, as shown in [\[link\]](#). On January 1, 2002, the exchange rate was 133 yen/dollar. On January 1, 2005, it was 103 yen/dollar. On June 1, 2007, it was 122 yen/dollar, on January 1, 2012, it was 77 yen per dollar, and on March 1, 2015, it was 120 yen per dollar. As investor sentiment swings back and forth, driving exchange rates up and down, exporters, importers, and banks involved in international lending are all affected. At worst, large movements in exchange rates can drive companies into bankruptcy or trigger a nationwide banking collapse. But even in the moderate case of the yen/dollar exchange rate, these movements of roughly 30 percent back and forth impose stress on both economies as firms must alter their export and import plans to take the new exchange rates into account. Especially in smaller countries where international trade is a relatively large share of GDP, exchange rate movements can rattle their economies.

U.S. Dollar Exchange Rate in Japanese Yen



Even seemingly stable exchange rates such as the Japanese Yen to the U.S. Dollar can vary when closely looked at over time. This figure

shows a relatively stable rate between 2011 and 2013. In 2013, there was a drastic depreciation of the Yen (relative to the U.S. Dollar) by about 14% and again at the end of the year in 2014 also by about 14%.

(Source: Federal Reserve Economic Data (FRED)
<https://research.stlouisfed.org/fred2/series/DEXJPUS>)

However, movements of floating exchange rates have advantages, too. After all, prices of goods and services rise and fall throughout a market economy, as demand and supply shift. If an economy experiences strong inflows or outflows of international financial capital, or has relatively high inflation, or if it experiences strong productivity growth so that purchasing power changes relative to other economies, then it makes economic sense for the exchange rate to shift as well.

Floating exchange rate advocates often argue that if government policies were more predictable and stable, then inflation rates and interest rates would be more predictable and stable. Exchange rates would bounce around less, too. The economist Milton Friedman (1912–2006), for example, wrote a defense of floating exchange rates in 1962 in his book *Capitalism and Freedom*:

"Being in favor of floating exchange rates does not mean being in favor of unstable exchange rates. When we support a free price system [for goods and services] at home, this does not imply that we favor a system in which prices fluctuate wildly up and down. What we want is a system in which prices are free to fluctuate but in which the forces determining them are sufficiently stable so that in fact prices move within moderate ranges. This is equally true in a system of floating exchange rates. The ultimate objective is a world in which exchange rates, while free to vary, are, in fact, highly stable because basic economic policies and conditions are stable."

Advocates of floating exchange rates admit that, yes, exchange rates may sometimes fluctuate. They point out, however, that if a central bank focuses on preventing either high inflation or deep recession, with low and

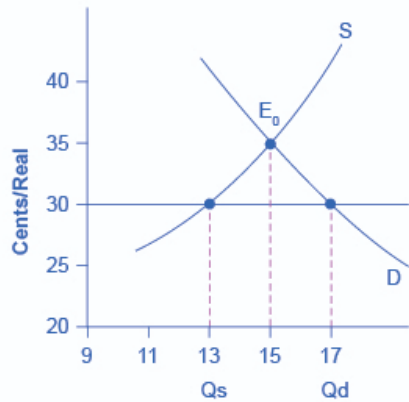
reasonably steady interest rates, then exchange rates will have less reason to vary.

Using Soft Pegs and Hard Pegs

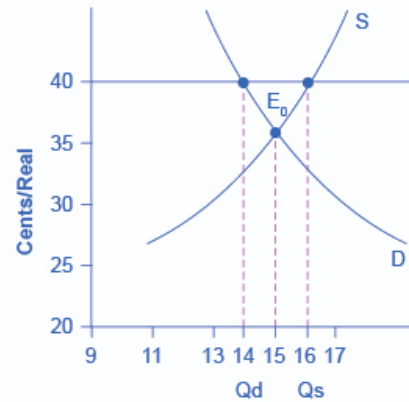
When a government intervenes in the foreign exchange market so that the exchange rate of its currency is different from what the market would have produced, it is said to have established a “peg” for its currency. A **soft peg** is the name for an exchange rate policy where the government usually allows the exchange rate to be set by the market, but in some cases, especially if the exchange rate seems to be moving rapidly in one direction, the central bank will intervene in the market. With a **hard peg** exchange rate policy, the central bank sets a fixed and unchanging value for the exchange rate. A central bank can implement soft peg and hard peg policies.

Suppose the market exchange rate for the Brazilian currency, the real, would be 35 cents/real with a daily quantity of 15 billion real traded in the market, as shown at the equilibrium E_0 in [\[link\]](#) (a) and [\[link\]](#) (b). However, the government of Brazil decides that the exchange rate should be 30 cents/real, as shown in [\[link\]](#) (a). Perhaps Brazil sets this lower exchange rate to benefit its export industries. Perhaps it is an attempt to stimulate aggregate demand by stimulating exports. Perhaps Brazil believes that the current market exchange rate is higher than the long-term purchasing power parity value of the real, so it is minimizing fluctuations in the real by keeping it at this lower rate. Perhaps the target exchange rate was set sometime in the past, and is now being maintained for the sake of stability. Whatever the reason, if Brazil’s central bank wishes to keep the exchange rate below the market level, it must face the reality that at this weaker exchange rate of 30 cents/real, the quantity demanded of its currency at 17 billion reals is greater than the quantity supplied of 13 billion reals in the foreign exchange market.

Pegging an Exchange Rate



(a) Pegging an exchange rate below equilibrium



(b) Pegging an exchange rate above equilibrium

(a) If an exchange rate is pegged below what would otherwise be the equilibrium, then the quantity demanded of the currency will exceed the quantity supplied. (b) If an exchange rate is pegged above what would otherwise be the equilibrium, then the quantity supplied of the currency exceeds the quantity demanded.

The Brazilian central bank could weaken its exchange rate in two ways. One approach is to use an expansionary monetary policy that leads to lower interest rates. In foreign exchange markets, the lower interest rates will reduce demand and increase supply of the real and lead to depreciation. This technique is not often used because lowering interest rates to weaken the currency may be in conflict with the country's monetary policy goals. Alternatively, Brazil's central bank could trade directly in the foreign exchange market. The central bank can expand the money supply by creating reals, use the reals to purchase foreign currencies, and avoid selling any of its own currency. In this way, it can fill the gap between quantity demanded and quantity supplied of its currency.

[\[link\]](#) (b) shows the opposite situation. Here, the Brazilian government desires a stronger exchange rate of 40 cents/real than the market rate of 35 cents/real. Perhaps Brazil desires the stronger currency to reduce aggregate demand and to fight inflation, or perhaps Brazil believes that that current market exchange rate is temporarily lower than the long-term rate.

Whatever the reason, at the higher desired exchange rate, the quantity supplied of 16 billion reals exceeds the quantity demanded of 14 billion reals.

Brazil's central bank can use a contractionary monetary policy to raise interest rates, which will increase demand and reduce supply of the currency on foreign exchange markets, and lead to an appreciation. Alternatively, Brazil's central bank can trade directly in the foreign exchange market. In this case, with an excess supply of its own currency in foreign exchange markets, the central bank must use reserves of foreign currency, like U.S. dollars, to demand its own currency and thus cause an appreciation of its exchange rate.

Both a soft peg and a hard peg policy require that the central bank intervene in the foreign exchange market. However, a hard peg policy attempts to preserve a fixed exchange rate at all times. A soft peg policy typically allows the exchange rate to move up and down by relatively small amounts in the short run of several months or a year, and to move by larger amounts over time, but seeks to avoid extreme short-term fluctuations.

Tradeoffs of Soft Pegs and Hard Pegs

When a country decides to alter the market exchange rate, it faces a number of tradeoffs. If it uses monetary policy to alter the exchange rate, it then cannot at the same time use monetary policy to address issues of inflation or recession. If it uses direct purchases and sales of foreign currencies in exchange rates, then it must face the issue of how it will handle its reserves of foreign currency. Finally, a pegged exchange rate can even create additional movements of the exchange rate; for example, even the possibility of government intervention in exchange rate markets will lead to rumors about whether and when the government will intervene, and dealers in the foreign exchange market will react to those rumors. Let's consider these issues in turn.

One concern with pegged exchange rate policies is that they imply a country's monetary policy is no longer focused on controlling inflation or shortening recessions, but now must also take the exchange rate into

account. For example, when a country pegs its exchange rate, it will sometimes face economic situations where it would like to have an expansionary monetary policy to fight recession—but it cannot do so because that policy would depreciate its exchange rate and break its hard peg. With a soft peg exchange rate policy, the central bank can sometimes ignore the exchange rate and focus on domestic inflation or recession—but in other cases the central bank may ignore inflation or recession and instead focus on its soft peg exchange rate. With a hard peg policy, domestic monetary policy is effectively no longer determined by domestic inflation or unemployment, but only by what monetary policy is needed to keep the exchange rate at the hard peg.

Another issue arises when a central bank intervenes directly in the exchange rate market. If a central bank ends up in a situation where it is perpetually creating and selling its own currency on foreign exchange markets, it will be buying the currency of other countries, like U.S. dollars or euros, to hold as reserves. Holding large reserves of other currencies has an opportunity cost, and central banks will not wish to boost such reserves without limit.

In addition, a central bank that causes a large increase in the supply of money is also risking an inflationary surge in aggregate demand. Conversely, when a central bank wishes to buy its own currency, it can do so by using its reserves of international currency like the U.S. dollar or the euro. But if the central bank runs out of such reserves, it can no longer use this method to strengthen its currency. Thus, buying foreign currencies in exchange rate markets can be expensive and inflationary, while selling foreign currencies can work only until a central bank runs out of reserves.

Yet another issue is that when a government pegs its exchange rate, it may unintentionally create another reason for additional fluctuation. With a soft peg policy, foreign exchange dealers and international investors react to every rumor about how or when the central bank is likely to intervene to influence the exchange rate, and as they react to rumors the exchange rate will shift up and down. Thus, even though the goal of a soft peg policy is to reduce short-term fluctuations of the exchange rate, the existence of the policy—when anticipated in the foreign exchange market—may sometimes

increase short-term fluctuations as international investors try to anticipate how and when the central bank will act.

A hard peg exchange rate policy will not allow short-term fluctuations in the exchange rate. If the government first announces a hard peg and then later changes its mind—perhaps the government becomes unwilling to keep interest rates high or to hold high levels of foreign exchange reserves—then the result of abandoning a hard peg could be a dramatic shift in the exchange rate.

In the mid-2000s, about one-third of the countries in the world used a soft peg approach and about one-quarter used a hard peg approach. The general trend in the 1990s was to shift away from a soft peg approach in favor of either floating rates or a hard peg. The concern is that a successful soft peg policy may, for a time, lead to very little variation in exchange rates, so that firms and banks in the economy begin to act as if a hard peg exists. When the exchange rate does move, the effects are especially painful because firms and banks have not planned and hedged against a possible change. Thus, the argument went, it is better either to be clear that the exchange rate is always flexible, or that it is fixed, but choosing an in-between soft peg option may end up being worst of all.

A Merged Currency

A final approach to exchange rate policy is for a nation to choose a common currency shared with one or more nations is also called a **merged currency**. A merged currency approach eliminates foreign exchange risk altogether. Just as no one worries about exchange rate movements when buying and selling between New York and California, Europeans know that the value of the euro will be the same in Germany and France and other European nations that have adopted the euro.

However, a merged currency also poses problems. Like a hard peg, a merged currency means that a nation has given up altogether on domestic monetary policy, and instead has put its interest rate policies in other hands. When Ecuador uses the U.S. dollar as its currency, it has no voice in whether the Federal Reserve raises or lowers interest rates. The European

Central Bank that determines monetary policy for the euro has representatives from all the euro nations. However, from the standpoint of, say, Portugal, there will be times when the decisions of the European Central Bank about monetary policy do not match the decisions that would have been made by a Portuguese central bank.

The lines between these four different exchange rate policies can blend into each other. For example, a soft peg exchange rate policy in which the government almost never acts to intervene in the exchange rate market will look a great deal like a floating exchange rate. Conversely, a soft peg policy in which the government intervenes often to keep the exchange rate near a specific level will look a lot like a hard peg. A decision to merge currencies with another country is, in effect, a decision to have a permanently fixed exchange rate with those countries, which is like a very hard exchange rate peg.

Global macroeconomics would be easier if the whole world had one currency and one central bank. The exchange rates between different currencies complicate the picture. If exchange rates are set solely by financial markets, they fluctuate substantially as short-term portfolio investors try to anticipate tomorrow's news. If the government attempts to intervene in exchange rate markets through soft pegs or hard pegs, it gives up at least some of the power to use monetary policy to focus on domestic inflations and recessions, and it risks causing even greater fluctuations in foreign exchange markets.

There is no consensus among economists about which exchange rate policies are best: floating, soft peg, hard peg, or merged currencies. The choice depends both on how well a nation's central bank can implement a specific exchange rate policy and on how well a nation's firms and banks can adapt to different exchange rate policies. A national economy that does a fairly good job at achieving the four main economic goals of growth, low inflation, low unemployment, and a sustainable balance of trade will probably do just fine most of the time with any exchange rate policy; conversely, no exchange rate policy is likely to save an economy that consistently fails at achieving these goals. On the other hand, a merged currency applied across wide geographic and cultural areas carries with it

its own set of problems, such as the ability for countries to conduct their own independent monetary policies.

Note:

Is a Stronger Dollar Good for the U.S. Economy?

The foreign exchange value of the dollar is a price and whether a higher price is good or bad depends on where you are standing: sellers benefit from higher prices and buyers are harmed. A stronger dollar is good for U.S. imports (and people working for U.S. importers) and U.S. investment abroad. It is also good for U.S. tourists going to other countries, since their dollar goes further. But a stronger dollar is bad for U.S. exports (and people working in U.S. export industries); it is bad for foreign investment in the United States (leading, for example, to higher U.S. interest rates); and it is bad for foreign tourists (as well as U.S. hotels, restaurants, and others in the tourist industry). In short, whether the U.S. dollar is good or bad is a more complex question than you may have thought. The economic answer is “it depends.”

Key Concepts and Summary

In a floating exchange rate policy, a country's exchange rate is determined in the foreign exchange market. In a soft peg exchange rate policy, a country's exchange rate is usually determined in the foreign exchange market, but the government sometimes intervenes to strengthen or weaken the exchange rate. In a hard peg exchange rate policy, the government chooses an exchange rate. A central bank can intervene in exchange markets in two ways. It can raise or lower interest rates to make the currency stronger or weaker. Or it can directly purchase or sell its currency in foreign exchange markets. All exchange rates policies face tradeoffs. A hard peg exchange rate policy will reduce exchange rate fluctuations, but means that a country must focus its monetary policy on the exchange rate, not on fighting recession or controlling inflation. When a nation merges its currency with another nation, it gives up on nationally oriented monetary policy altogether.

A soft peg exchange rate may create additional volatility as exchange rate markets try to anticipate when and how the government will intervene. A flexible exchange rate policy allows monetary policy to focus on inflation and unemployment, and allows the exchange rate to change with inflation and rates of return, but also raises a risk that exchange rates may sometimes make large and abrupt movements. The spectrum of exchange rate policies includes: (a) a floating exchange rate, (b) a pegged exchange rate, soft or hard, and (c) a merged currency. Monetary policy can focus on a variety of goals: (a) inflation; (b) inflation or unemployment, depending on which is the most dangerous obstacle; and (c) a long-term rule based policy designed to keep the money supply stable and predictable.

Self-Check Questions

Exercise:

Problem:

How would a contractionary monetary policy affect the exchange rate, net exports, aggregate demand, and aggregate supply?

Solution:

A contractionary monetary policy, by driving up domestic interest rates, would cause the currency to appreciate. The higher value of the currency in foreign exchange markets would reduce exports, since from the perspective of foreign buyers, they are now more expensive. The higher value of the currency would similarly stimulate imports, since they would now be cheaper from the perspective of domestic buyers. Lower exports and higher imports cause net exports ($EX - IM$) to fall, which causes aggregate demand to fall. The result would be a decrease in GDP working through the exchange rate mechanism reinforcing the effect contractionary monetary policy has on domestic investment expenditure. However, cheaper imports would stimulate aggregate supply, bringing GDP back to potential, though at a lower price level.

Exercise:

Problem:

A central bank can allow its currency to fall indefinitely, but it cannot allow its currency to rise indefinitely. Why not?

Solution:

For a currency to fall, a central bank need only supply more of its currency in foreign exchange markets. It can print as much domestic currency as it likes. For a currency to rise, a central bank needs to buy its currency in foreign exchange markets, paying with foreign currency. Since no central bank has an infinite amount of foreign currency reserves, it cannot buy its currency indefinitely.

Exercise:**Problem:**

Is a country for which imports and exports make up a large fraction of the GDP more likely to adopt a flexible exchange rate or a fixed (hard peg) exchange rate?

Solution:

Variations in exchange rates, because they change import and export prices, disturb international trade flows. When trade is a large part of a nation's economic activity, government will find it more advantageous to fix exchange rates to minimize disruptions of trade flows.

Review Questions**Exercise:****Problem:**

What is the difference between a floating exchange rate, a soft peg, a hard peg, and dollarization?

Exercise:

Problem:

List some advantages and disadvantages of the different exchange rate policies.

Critical Thinking Questions**Exercise:****Problem:**

Many developing countries, like Mexico, have moderate to high rates of inflation. At the same time, international trade plays an important role in their economies. What type of exchange rate regime would be best for such a country's currency *vis à vis* the U.S. dollar?

Exercise:**Problem:**

What would make a country decide to change from a common currency, like the euro, back to its own currency?

References

Friedman, Milton. *Capitalism and Freedom*. Chicago: University of Chicago Press, 1962.

Glossary

floating exchange rate

a country lets the value of its currency be determined in the exchange rate market

hard peg

an exchange rate policy in which the central bank sets a fixed and unchanging value for the exchange rate

international capital flows

flow of financial capital across national boundaries either as portfolio investment or direct investment

merged currency

when a nation chooses to use the currency of another nation

soft peg

an exchange rate policy in which the government usually allows the exchange rate to be set by the market, but in some cases, especially if the exchange rate seems to be moving rapidly in one direction, the central bank will intervene

Tobin taxes

see international capital flows

Introduction to International Trade

class="introduction"

Apple or Samsung iPhone?

While the iPhone is readily recognized as an Apple product, 26% of the component costs in it come from components made by rival phone-maker, Samsung. In international trade, there are often “conflicts” like this as each country or company focuses on what it does best.

(Credit: modification of work by Yutaka Tsutano
Creative Commons)

**Note:****Just Whose iPhone Is It?**

The iPhone is a global product. Apple does not manufacture the iPhone components, nor does it assemble them. The assembly is done by Foxconn Corporation, a Taiwanese company, at its factory in Sengzhen, China. But, Samsung, the electronics firm and competitor to Apple, actually supplies many of the parts that make up an iPhone—about 26%. That means, that Samsung is both the biggest supplier and biggest competitor for Apple. Why do these two firms work together to produce the iPhone? To understand the economic logic behind international trade, you have to accept, as these firms do, that trade is about mutually beneficial exchange. Samsung is one of the world's largest electronics parts suppliers. Apple lets Samsung focus on making the best parts, which allows Apple to concentrate on its strength—designing elegant products that are easy to use. If each company (and by extension each country) focuses on what it does best, there will be gains for all through trade.

Note:**Introduction to International Trade**

In this chapter, you will learn about:

- Absolute and Comparative Advantage
- What Happens When a Country Has an Absolute Advantage in All Goods
- Intra-industry Trade between Similar Economies
- The Benefits of Reducing Barriers to International Trade

We live in a global marketplace. The food on your table might include fresh fruit from Chile, cheese from France, and bottled water from Scotland. Your wireless phone might have been made in Taiwan or Korea. The clothes you wear might be designed in Italy and manufactured in China. The toys you give to a child might have come from India. The car you drive might come from Japan, Germany, or Korea. The gasoline in the tank might be refined from crude oil from Saudi Arabia, Mexico, or Nigeria. As a worker, if your job is involved with farming, machinery, airplanes, cars, scientific instruments, or many other technology-related industries, the odds are good that a hearty proportion of the sales of your employer—and hence the money that pays your salary—comes from export sales. We are all linked by international trade, and the volume of that trade has grown dramatically in the last few decades.

The first wave of globalization started in the nineteenth century and lasted up to the beginning of World War I. Over that time, global exports as a share of global GDP rose from less than 1% of GDP in 1820 to 9% of GDP in 1913. As the Nobel Prize-winning economist Paul Krugman of Princeton University wrote in 1995:

"It is a late-twentieth-century conceit that we invented the global economy just yesterday. In fact, world markets achieved an impressive degree of integration during the second half of the nineteenth century. Indeed, if one wants a specific date for the beginning of a truly global economy, one might well choose 1869, the year in which both the Suez Canal and the Union

Pacific railroad were completed. By the eve of the First World War steamships and railroads had created markets for standardized commodities, like wheat and wool, that were fully global in their reach. Even the global flow of information was better than modern observers, focused on electronic technology, tend to realize: the first submarine telegraph cable was laid under the Atlantic in 1858, and by 1900 all of the world's major economic regions could effectively communicate instantaneously."

This first wave of globalization crashed to a halt in the beginning of the twentieth century. World War I severed many economic connections. During the Great Depression of the 1930s, many nations misguidedly tried to fix their own economies by reducing foreign trade with others. World War II further hindered international trade. Global flows of goods and financial capital rebuilt themselves only slowly after World War II. It was not until the early 1980s that global economic forces again became as important, relative to the size of the world economy, as they were before World War I.

Absolute and Comparative Advantage

By the end of this section, you will be able to:

- Define absolute advantage, comparative advantage, and opportunity costs
- Explain the gains of trade created when a country specializes

The American statesman Benjamin Franklin (1706–1790) once wrote: “No nation was ever ruined by trade.” Many economists would express their attitudes toward international trade in an even more positive manner. The evidence that international trade confers overall benefits on economies is pretty strong. Trade has accompanied economic growth in the United States and around the world. Many of the national economies that have shown the most rapid growth in the last few decades—for example, Japan, South Korea, China, and India—have done so by dramatically orienting their economies toward international trade. There is no modern example of a country that has shut itself off from world trade and yet prospered. To understand the benefits of trade, or why we trade in the first place, we need to understand the concepts of comparative and absolute advantage.

In 1817, David Ricardo, a businessman, economist, and member of the British Parliament, wrote a treatise called *On the Principles of Political Economy and Taxation*. In this treatise, Ricardo argued that specialization and free trade benefit all trading partners, even those that may be relatively inefficient. To see what he meant, we must be able to distinguish between absolute and comparative advantage.

A country has an **absolute advantage** in producing a good over another country if it uses fewer resources to produce that good. Absolute advantage can be the result of a country’s natural endowment. For example, extracting oil in Saudi Arabia is pretty much just a matter of “drilling a hole.” Producing oil in other countries can require considerable exploration and costly technologies for drilling and extraction—if indeed they have any oil at all. The United States has some of the richest farmland in the world, making it easier to grow corn and wheat than in many other countries. Guatemala and Colombia have climates especially suited for growing coffee. Chile and Zambia have some of the world’s richest copper mines.

As some have argued, “geography is destiny.” Chile will provide copper and Guatemala will produce coffee, and they will trade. When each country has a product others need and it can be produced with fewer resources in one country over another, then it is easy to imagine all parties benefitting from trade. However, thinking about trade just in terms of geography and absolute advantage is incomplete. Trade really occurs because of comparative advantage.

Recall from the chapter [Choice in a World of Scarcity](#) that a country has a comparative advantage when a good can be produced at a lower cost in terms of other goods. The question each country or company should be asking when it trades is this: “What do we give up to produce this good?” It should be no surprise that the concept of comparative advantage is based on this idea of opportunity cost from [Choice in a World of Scarcity](#). For example, if Zambia focuses its resources on producing copper, its labor, land and financial resources cannot be used to produce other goods such as corn. As a result, Zambia gives up the opportunity to produce corn. How do we quantify the cost in terms of other goods? Simplify the problem and assume that Zambia just needs labor to produce copper and corn. The companies that produce either copper or corn tell you that it takes 10 hours to mine a ton of copper and 20 hours to harvest a bushel of corn. This means the opportunity cost of producing a ton of copper is 2 bushels of corn. The next section develops absolute and comparative advantage in greater detail and relates them to trade.

Note:

Visit this [website](#) for a list of articles and podcasts pertaining to international trade topics.



A Numerical Example of Absolute and Comparative Advantage

Consider a hypothetical world with two countries, Saudi Arabia and the United States, and two products, oil and corn. Further assume that consumers in both countries desire both these goods. These goods are homogeneous, meaning that consumers/producers cannot differentiate between corn or oil from either country. There is only one resource available in both countries, labor hours. Saudi Arabia can produce oil with fewer resources, while the United States can produce corn with fewer resources. [\[link\]](#) illustrates the advantages of the two countries, expressed in terms of how many hours it takes to produce one unit of each good.

Country	Oil (hours per barrel)	Corn (hours per bushel)
Saudi Arabia	1	4
United States	2	1

How Many Hours It Takes to Produce Oil and Corn

In [\[link\]](#), Saudi Arabia has an absolute advantage in the production of oil because it only takes an hour to produce a barrel of oil compared to two hours in the United States. The United States has an absolute advantage in the production of corn.

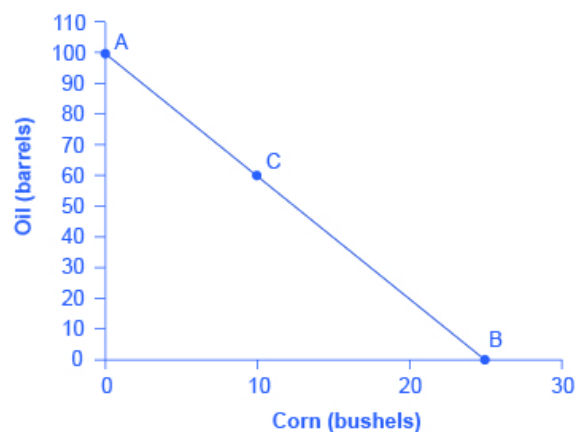
To simplify, let's say that Saudi Arabia and the United States each have 100 worker hours (see [\[link\]](#)). We illustrate what each country is capable of producing on its own using a production possibility frontier (PPF) graph,

shown in [\[link\]](#). Recall from [Choice in a World of Scarcity](#) that the production possibilities frontier shows the maximum amount that each country can produce given its limited resources, in this case workers, and its level of technology.

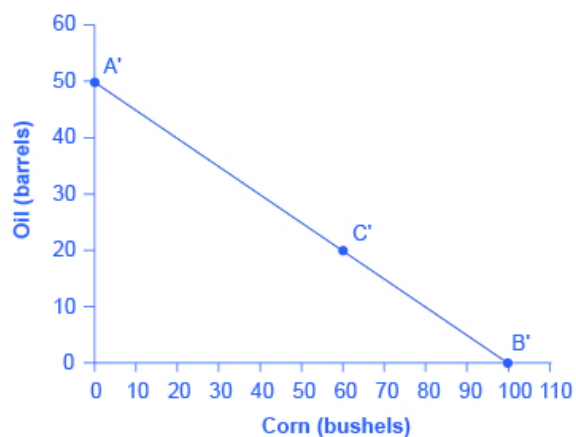
Country	Oil Production using 100 worker hours (barrels)		Corn Production using 100 worker hours (bushels)
Saudi Arabia	100	or	25
United States	50	or	100

Production Possibilities before Trade

Production Possibilities Frontiers



(a) Saudi Arabia



(b) The United States

(a) Saudi Arabia can produce 100 barrels of oil at maximum and zero corn (point A), or 25 bushels of corn and zero oil (point B). It can also produce other combinations of oil and corn if it wants to consume both

goods, such as at point C. Here it chooses to produce/consume 60 barrels of oil, leaving 40 work hours that can be allocated to producing 10 bushels of corn, using the data in [\[link\]](#). (b) If the United States produces only oil, it can produce, at maximum, 50 barrels and zero corn (point A'), or at the other extreme, it can produce a maximum of 100 bushels of corn and no oil (point B'). Other combinations of both oil and corn are possible, such as point C'. All points above the frontiers are impossible to produce given the current level of resources and technology.

Arguably Saudi and U.S. consumers desire both oil and corn to live. Let's say that before trade occurs, both countries produce and consume at point C or C'. Thus, before trade, the Saudi Arabian economy will devote 60 worker hours to produce oil, as shown in [\[link\]](#). Given the information in [\[link\]](#), this choice implies that it produces/consumes 60 barrels of oil. With the remaining 40 worker hours, since it needs four hours to produce a bushel of corn, it can produce only 10 bushels. To be at point C', the U.S. economy devotes 40 worker hours to produce 20 barrels of oil and the remaining worker hours can be allocated to produce 60 bushels of corn.

Country	Oil Production (barrels)	Corn Production (bushels)
Saudi Arabia (C)	60	10
United States (C')	20	60
Total World Production	80	70

Production before Trade

The slope of the production possibility frontier illustrates the opportunity cost of producing oil in terms of corn. Using all its resources, the United States can produce 50 barrels of oil *or* 100 bushels of corn. So the opportunity cost of one barrel of oil is two bushels of corn—or the slope is $1/2$. Thus, in the U.S. production possibility frontier graph, every increase in oil production of one barrel implies a decrease of two bushels of corn. Saudi Arabia can produce 100 barrels of oil *or* 25 bushels of corn. The opportunity cost of producing one barrel of oil is the loss of $1/4$ of a bushel of corn that Saudi workers could otherwise have produced. In terms of corn, notice that Saudi Arabia gives up the least to produce a barrel of oil. These calculations are summarized in [\[link\]](#).

Country	Opportunity cost of one unit — Oil (in terms of corn)	Opportunity cost of one unit — Corn (in terms of oil)
Saudi Arabia	$\frac{1}{4}$	4
United States	2	$\frac{1}{2}$

Opportunity Cost and Comparative Advantage

Again recall that comparative advantage was defined as the opportunity cost of producing goods. Since Saudi Arabia gives up the least to produce a barrel of oil, ($\frac{1}{4} < 2$ in [\[link\]](#)) it has a comparative advantage in oil production. The United States gives up the least to produce a bushel of corn, so it has a comparative advantage in corn production.

In this example, there is symmetry between absolute and comparative advantage. Saudi Arabia needs fewer worker hours to produce oil (absolute advantage, see [\[link\]](#)), and also gives up the least in terms of other goods to produce oil (comparative advantage, see [\[link\]](#)). Such symmetry is not always the case, as we will show after we have discussed gains from trade fully. But first, read the following Clear It Up feature to make sure you understand why the PPF line in the graphs is straight.

Note:

Can a production possibility frontier be straight?

When you first met the production possibility frontier (PPF) in the chapter on [Choice in a World of Scarcity](#), it was drawn with an outward-bending shape. This shape illustrated that as inputs were transferred from producing one good to another—like from education to health services—there were increasing opportunity costs. In the examples in this chapter, the PPFs are drawn as straight lines, which means that opportunity costs are constant. When a marginal unit of labor is transferred away from growing corn and toward producing oil, the decline in the quantity of corn and the increase in the quantity of oil is always the same. In reality this is possible only if the contribution of additional workers to output did not change as the scale of production changed. The linear production possibilities frontier is a less realistic model, but a straight line simplifies calculations. It also illustrates economic themes like absolute and comparative advantage just as clearly.

Gains from Trade

Consider the trading positions of the United States and Saudi Arabia after they have specialized and traded. Before trade, Saudi Arabia produces/consumes 60 barrels of oil and 10 bushels of corn. The United States produces/consumes 20 barrels of oil and 60 bushels of corn. Given their current production levels, if the United States can trade an amount of corn fewer than 60 bushels and receives in exchange an amount of oil greater than 20 barrels, it will **gain from trade**. With trade, the United States can consume more of both goods than it did without specialization

and trade. (Recall that the chapter [Welcome to Economics!](#) defined specialization as it applies to workers and firms. Specialization is also used to describe the occurrence when a country shifts resources to focus on producing a good that offers comparative advantage.) Similarly, if Saudi Arabia can trade an amount of oil less than 60 barrels and receive in exchange an amount of corn greater than 10 bushels, it will have more of both goods than it did before specialization and trade. [\[link\]](#) illustrates the range of trades that would benefit both sides.

The U.S. Economy, after Specialization, Will Benefit If It:	The Saudi Arabian Economy, after Specialization, Will Benefit If It:
Exports no more than 60 bushels of corn	Imports at least 10 bushels of corn
Imports at least 20 barrels of oil	Exports less than 60 barrels of oil

The Range of Trades That Benefit Both the United States and Saudi Arabia

The underlying reason why trade benefits both sides is rooted in the concept of opportunity cost, as the following Clear It Up feature explains. If Saudi Arabia wishes to expand domestic production of corn in a world without international trade, then based on its opportunity costs it must give up four barrels of oil for every one additional bushel of corn. If Saudi Arabia could find a way to give up less than four barrels of oil for an additional bushel of corn (or equivalently, to receive more than one bushel of corn for four barrels of oil), it would be better off.

Note:

What are the opportunity costs and gains from trade?

The range of trades that will benefit each country is based on the country's opportunity cost of producing each good. The United States can produce 100 bushels of corn or 50 barrels of oil. For the United States, the opportunity cost of producing one barrel of oil is two bushels of corn. If we divide the numbers above by 50, we get the same ratio: one barrel of oil is equivalent to two bushels of corn, or $(100/50 = 2)$ and $(50/50 = 1)$. In a trade with Saudi Arabia, if the United States is going to give up 100 bushels of corn in exports, it must import at least 50 barrels of oil to be just as well off. Clearly, to gain from trade it needs to be able to gain more than a half barrel of oil for its bushel of corn—or why trade at all?

Recall that David Ricardo argued that if each country specializes in its comparative advantage, it will benefit from trade, and total global output will increase. How can we show gains from trade as a result of comparative advantage and specialization? [\[link\]](#) shows the output assuming that each country specializes in its comparative advantage and produces no other good. This is 100% specialization. Specialization leads to an increase in total world production. (Compare the total world production in [\[link\]](#) to that in [\[link\]](#).)

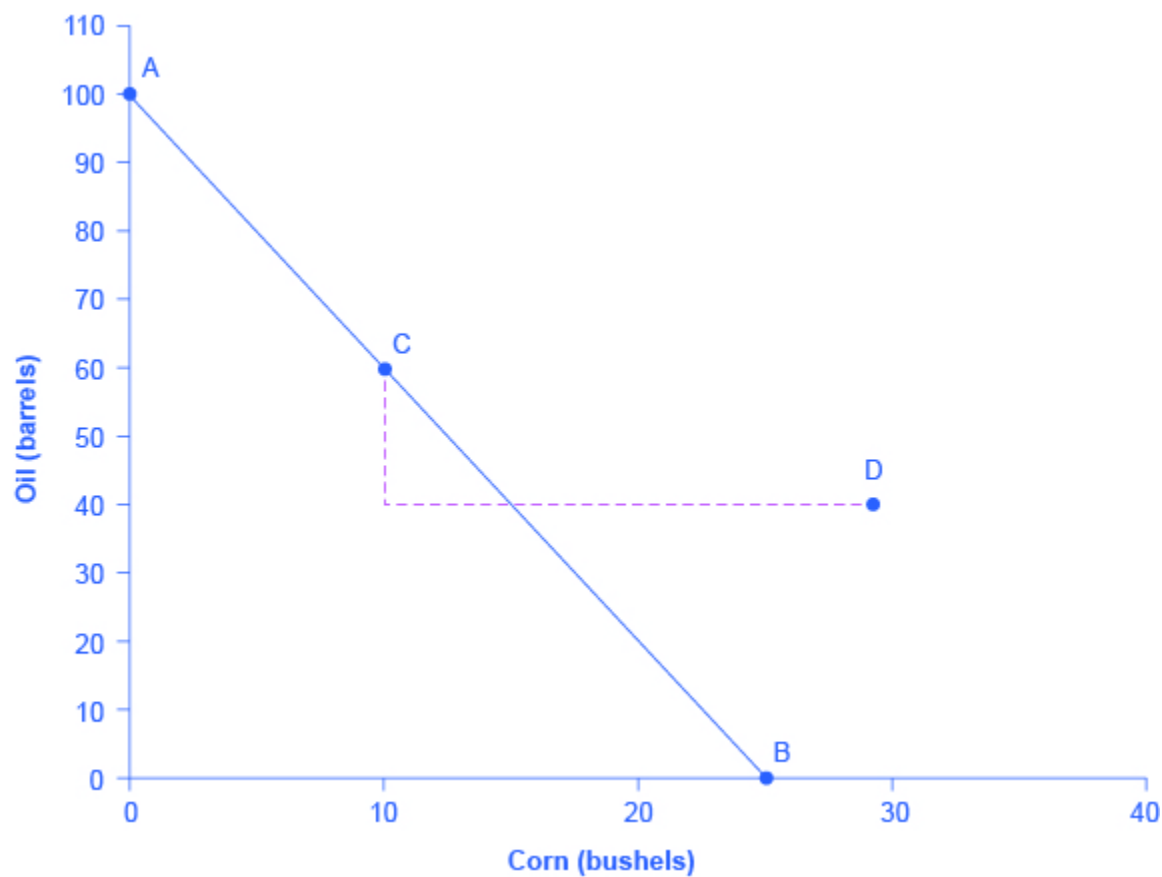
Country	Quantity produced after 100% specialization — Oil (barrels)	Quantity produced after 100% specialization — Corn (bushels)
Saudi Arabia	100	0

Country	Quantity produced after 100% specialization — Oil (barrels)	Quantity produced after 100% specialization — Corn (bushels)
United States	0	100
Total World Production	100	100

How Specialization Expands Output

What if we did not have complete specialization, as in [\[link\]](#)? Would there still be gains from trade? Consider another example, such as when the United States and Saudi Arabia start at C and C', respectively, as shown in [\[link\]](#). Consider what occurs when trade is allowed and the United States exports 20 bushels of corn to Saudi Arabia in exchange for 20 barrels of oil.

Production Possibilities Frontier in Saudi Arabia



Gains from trade of oil can increase only by achieving less from trade of corn. The opposite is true as well: The more gains from trade of corn, the fewer gains from trade of oil.

Starting at point C, which shows Saudi oil production of 60, reduce Saudi Oil domestic oil consumption by 20, since 20 is exported to United States and exchanged for 20 units of corn. This enables Saudi to reach point D, where consumption of oil is now 40 barrels and consumption of corn has increased to 30 (see [\[link\]](#)). Notice that even without 100% specialization, if the “trading price,” in this case 20 barrels of oil for 20 bushels of corn, is greater than the country’s opportunity cost, the Saudis will gain from trade. Since the post-trade consumption point D is beyond its production possibility frontier, hence Saudi Arabia has gained from trade.

Note:

Visit this [website](#) for trade-related data visualizations.



Key Concepts and Summary

A country has an absolute advantage in those products in which it has a productivity edge over other countries; it takes fewer resources to produce a product. A country has a comparative advantage when a good can be produced at a lower cost in terms of other goods. Countries that specialize based on comparative advantage gain from trade.

Self-Check Questions

Exercise:**Problem:**

True or False: The source of comparative advantage must be natural elements like climate and mineral deposits. Explain.

Solution:

False. Anything that leads to different levels of productivity between two economies can be a source of comparative advantage. For example, the education of workers, the knowledge base of engineers and scientists in a country, the part of a split-up value chain where they have their specialized learning, economies of scale, and other factors can all determine comparative advantage.

Exercise:**Problem:**

Brazil can produce 100 pounds of beef or 10 autos; in contrast the United States can produce 40 pounds of beef or 30 autos. Which country has the absolute advantage in beef? Which country has the absolute advantage in producing autos? What is the opportunity cost of producing one pound of beef in Brazil? What is the opportunity cost of producing one pound of beef in the United States?

Solution:

Brazil has the absolute advantage in producing beef and the United States has the absolute advantage in autos. The opportunity cost of producing one pound of beef is $1/10$ of an auto; in the United States it is $3/4$ of an auto.

Exercise:**Problem:**

In France it takes one worker to produce one sweater, and one worker to produce one bottle of wine. In Tunisia it takes two workers to produce one sweater, and three workers to produce one bottle of wine. Who has the absolute advantage in production of sweaters? Who has the absolute advantage in the production of wine? How can you tell?

Solution:

In answering questions like these, it is often helpful to begin by organizing the information in a table, such as in the following table. Notice that, in this case, the productivity of the countries is expressed in terms of how many workers it takes to produce a unit of a product.

Country	One Sweater	One Bottle of wine
France	1 worker	1 worker
Tunisia	2 workers	3 workers

In this example, France has an absolute advantage in the production of both sweaters and wine. You can tell because it takes France less labor to produce a unit of the good.

Review Questions

Exercise:

Problem:

What is absolute advantage? What is comparative advantage?

Exercise:

Problem:

Under what conditions does comparative advantage lead to gains from trade?

Exercise:

Problem:

What factors does Paul Krugman identify that supported the expansion of international trade in the 1800s?

Critical Thinking Questions

Exercise:

Problem:

Are differences in geography behind the differences in absolute advantages?

Exercise:**Problem:**

Why does the United States not have an absolute advantage in coffee?

Exercise:**Problem:**

Look at [\[link\]](#). Compute the opportunity costs of producing sweaters and wine in both France and Tunisia. Who has the lowest opportunity cost of producing sweaters and who has the lowest opportunity cost of producing wine? Explain what it means to have a lower opportunity cost.

Problems**Exercise:****Problem:**

France and Tunisia both have Mediterranean climates that are excellent for producing/harvesting green beans and tomatoes. In France it takes two hours for each worker to harvest green beans and two hours to harvest a tomato. Tunisian workers need only one hour to harvest the tomatoes but four hours to harvest green beans. Assume there are only two workers, one in each country, and each works 40 hours a week.

- a. Draw a production possibilities frontier for each country. *Hint:* Remember the production possibility frontier is the maximum that all workers can produce at a unit of time which, in this problem, is a week.

- b. Identify which country has the absolute advantage in green beans and which country has the absolute advantage in tomatoes.
- c. Identify which country has the comparative advantage.
- d. How much would France have to give up in terms of tomatoes to gain from trade? How much would it have to give up in terms of green beans?

References

Krugman, Paul R. *Pop Internationalism*. The MIT Press, Cambridge. 1996.

Krugman, Paul R. "What Do Undergrads Need to Know about Trade?" *American Economic Review* 83, no. 2. 1993. 23-26.

Ricardo, David. *On the Principles of Political Economy and Taxation*. London: John Murray, 1817.

Ricardo, David. "On the Principles of Political Economy and Taxation." *Library of Economics and Liberty*.
<http://www.econlib.org/library/Ricardo/ricP.html>.

Glossary

absolute advantage

when one country can use fewer resources to produce a good compared to another country; when a country is more productive compared to another country

gain from trade

a country that can consume more than it can produce as a result of specialization and trade

What Happens When a Country Has an Absolute Advantage in All Goods

By the end of this section, you will be able to:

- Show the relationship between production costs and comparative advantage
- Identify situations of mutually beneficial trade
- Identify trade benefits by considering opportunity costs

What happens to the possibilities for trade if one country has an absolute advantage in everything? This is typical for high-income countries that often have well-educated workers, technologically advanced equipment, and the most up-to-date production processes. These high-income countries can produce all products with fewer resources than a low-income country. If the high-income country is more productive across the board, will there still be gains from trade? Good students of Ricardo understand that trade is about mutually beneficial exchange. Even when one country has an absolute advantage in all products, trade can still benefit both sides. This is because gains from trade come from specializing in one's comparative advantage.

Production Possibilities and Comparative Advantage

Consider the example of trade between the United States and Mexico described in [\[link\]](#). In this example, it takes four U.S. workers to produce 1,000 pairs of shoes, but it takes five Mexican workers to do so. It takes one U.S. worker to produce 1,000 refrigerators, but it takes four Mexican workers to do so. The United States has an absolute advantage in productivity with regard to both shoes and refrigerators; that is, it takes fewer workers in the United States than in Mexico to produce both a given number of shoes and a given number of refrigerators.

Country	Number of Workers needed to produce 1,000 units — Shoes	Number of Workers needed to produce 1,000 units — Refrigerators
United States	4 workers	1 worker
Mexico	5 workers	4 workers

Resources Needed to Produce Shoes and Refrigerators

Absolute advantage simply compares the productivity of a worker between countries. It answers the question, “How many inputs do I need to produce shoes in Mexico?” Comparative advantage asks this same question slightly differently. Instead of comparing how many workers it takes to produce a good, it asks, “How much am I giving up to produce this good in this country?” Another way of looking at this is that comparative advantage identifies the good for which the producer’s absolute advantage is relatively larger, or where the producer’s absolute productivity disadvantage is relatively smaller. The United States can produce 1,000 shoes with four-fifths as many workers as Mexico (four versus five), but it can produce 1,000 refrigerators with only one-quarter as many workers (one versus four). So, the comparative advantage of the United States, where its absolute productivity advantage is relatively greatest, lies with refrigerators, and Mexico’s comparative advantage, where its absolute productivity disadvantage is least, is in the production of shoes.

Mutually Beneficial Trade with Comparative Advantage

When nations increase production in their area of comparative advantage and trade with each other, both countries can benefit. Again, the production possibility frontier is a useful tool to visualize this benefit.

Consider a situation where the United States and Mexico each have 40 workers. For example, as [\[link\]](#) shows, if the United States divides its labor so that 40 workers are making shoes, then, since it takes four workers in the

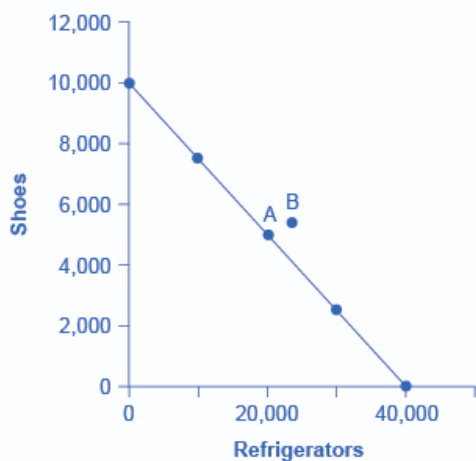
United States to make 1,000 shoes, a total of 10,000 shoes will be produced. (If four workers can make 1,000 shoes, then 40 workers will make 10,000 shoes). If the 40 workers in the United States are making refrigerators, and each worker can produce 1,000 refrigerators, then a total of 40,000 refrigerators will be produced.

Country	Shoe Production — using 40 workers		Refrigerator Production — using 40 workers
United States	10,000 shoes	or	40,000 refrigerators
Mexico	8,000 shoes	or	10,000 refrigerators

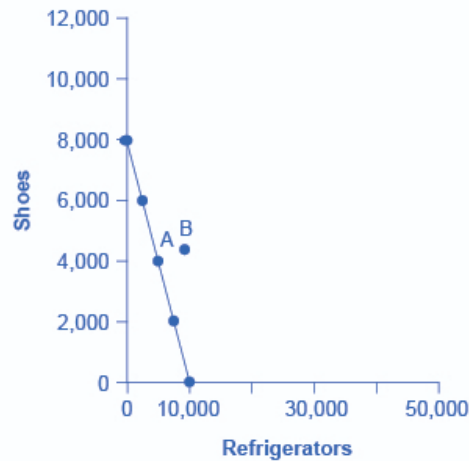
Production Possibilities before Trade with Complete Specialization

As always, the slope of the production possibility frontier for each country is the opportunity cost of one refrigerator in terms of foregone shoe production—when labor is transferred from producing the latter to producing the former (see [link](#)).

Production Possibility Frontiers



(a) U.S. PPF (40 workers)



(b) Mexico PPF (40 workers)

(a) With 40 workers, the United States can produce either 10,000 shoes and zero refrigerators or 40,000 refrigerators and zero shoes. (b) With 40 workers, Mexico can produce a maximum of 8,000 shoes and zero refrigerators, or 10,000 refrigerators and zero shoes. All other points on the production possibility line are possible combinations of the two goods that can be produced given current resources. Point A on both graphs is where the countries start producing and consuming before trade. Point B is where they end up after trade.

Let's say that, in the situation before trade, each nation prefers to produce a combination of shoes and refrigerators that is shown at point A. [\[link\]](#) shows the output of each good for each country and the total output for the two countries.

Country	Current Shoe Production	Current Refrigerator Production
---------	-------------------------	---------------------------------

Country	Current Shoe Production	Current Refrigerator Production
United States	5,000	20,000
Mexico	4,000	5,000
Total	9,000	25,000

Total Production at Point A before Trade

Continuing with this scenario, each country transfers some amount of labor toward its area of comparative advantage. For example, the United States transfers six workers away from shoes and toward producing refrigerators. As a result, U.S. production of shoes decreases by 1,500 units ($6/4 \times 1,000$), while its production of refrigerators increases by 6,000 (that is, $6/1 \times 1,000$). Mexico also moves production toward its area of comparative advantage, transferring 10 workers away from refrigerators and toward production of shoes. As a result, production of refrigerators in Mexico falls by 2,500 ($10/4 \times 1,000$), but production of shoes increases by 2,000 pairs ($10/5 \times 1,000$). Notice that when both countries shift production toward each of their comparative advantages (what they are relatively better at), their combined production of both goods rises, as shown in [\[link\]](#). The reduction of shoe production by 1,500 pairs in the United States is more than offset by the gain of 2,000 pairs of shoes in Mexico, while the reduction of 2,500 refrigerators in Mexico is more than offset by the additional 6,000 refrigerators produced in the United States.

Country	Shoe Production	Refrigerator Production
---------	-----------------	-------------------------

Country	Shoe Production	Refrigerator Production
United States	3,500	26,000
Mexico	6,000	2,500
Total	9,500	28,500

Shifting Production Toward Comparative Advantage Raises Total Output

This numerical example illustrates the remarkable insight of comparative advantage: even when one country has an absolute advantage in all goods and another country has an absolute disadvantage in all goods, both countries can still benefit from trade. Even though the United States has an absolute advantage in producing both refrigerators and shoes, it makes economic sense for it to specialize in the good for which it has a comparative advantage. The United States will export refrigerators and in return import shoes.

How Opportunity Cost Sets the Boundaries of Trade

This example shows that both parties can benefit from specializing in their comparative advantages and trading. By using the opportunity costs in this example, it is possible to identify the range of possible trades that would benefit each country.

Mexico started out, before specialization and trade, producing 4,000 pairs of shoes and 5,000 refrigerators (see [\[link\]](#) and [\[link\]](#)). Then, in the numerical example given, Mexico shifted production toward its comparative advantage and produced 6,000 pairs of shoes but only 2,500 refrigerators. Thus, if Mexico can export no more than 2,000 pairs of shoes (giving up 2,000 pairs of shoes) in exchange for imports of at least 2,500 refrigerators (a gain of 2,500 refrigerators), it will be able to consume more of both goods than before trade. Mexico will be unambiguously better off. Conversely, the United States started off, before specialization and trade, producing 5,000 pairs of shoes and 20,000 refrigerators. In the example, it

then shifted production toward its comparative advantage, producing only 3,500 shoes but 26,000 refrigerators. If the United States can export no more than 6,000 refrigerators in exchange for imports of at least 1,500 pairs of shoes, it will be able to consume more of both goods and will be unambiguously better off.

The range of trades that can benefit both nations is shown in [\[link\]](#). For example, a trade where the U.S. exports 4,000 refrigerators to Mexico in exchange for 1,800 pairs of shoes would benefit both sides, in the sense that both countries would be able to consume more of both goods than in a world without trade.

The U.S. economy, after specialization, will benefit if it:	The Mexican economy, after specialization, will benefit if it:
<i>Exports</i> fewer than 6,000 refrigerators	<i>Imports</i> at least 2,500 refrigerators
<i>Imports</i> at least 1,500 pairs of shoes	<i>Exports</i> no more than 2,000 pairs of shoes

The Range of Trades That Benefit Both the United States and Mexico

Trade allows each country to take advantage of lower opportunity costs in the other country. If Mexico wants to produce more refrigerators without trade, it must face its domestic opportunity costs and reduce shoe production. If Mexico, instead, produces more shoes and then trades for refrigerators made in the United States, where the opportunity cost of producing refrigerators is lower, Mexico can in effect take advantage of the lower opportunity cost of refrigerators in the United States. Conversely, when the United States specializes in its comparative advantage of refrigerator production and trades for shoes produced in Mexico,

international trade allows the United States to take advantage of the lower opportunity cost of shoe production in Mexico.

The theory of comparative advantage explains why countries trade: they have different comparative advantages. It shows that the gains from international trade result from pursuing comparative advantage and producing at a lower opportunity cost. The following Work It Out feature shows how to calculate absolute and comparative advantage and the way to apply them to a country’s production.

Note:
Calculating Absolute and Comparative Advantage
In Canada a worker can produce 20 barrels of oil or 40 tons of lumber. In Venezuela, a worker can produce 60 barrels of oil or 30 tons of lumber.

Country	Oil (barrels)		Lumber (tons)
Canada	20	or	40
Venezuela	60	or	30

a. Who has the absolute advantage in the production of oil or lumber? How can you tell?

b. Which country has a comparative advantage in the production of oil?

c. Which country has a comparative advantage in producing lumber?

d. In this example, is absolute advantage the same as comparative advantage, or not?

e. In what product should Canada specialize? In what product should Venezuela specialize?

Step 1. Make a table like [\[link\]](#).

Step 2. To calculate absolute advantage, look at the larger of the numbers for each product. One worker in Canada can produce more lumber (40 tons versus 30 tons), so Canada has the absolute advantage in lumber. One worker in Venezuela can produce 60 barrels of oil compared to a worker in Canada who can produce only 20.

Step 3. To calculate comparative advantage, find the opportunity cost of producing one barrel of oil in both countries. The country with the lowest opportunity cost has the comparative advantage. With the same labor time, Canada can produce either 20 barrels of oil or 40 tons of lumber. So in effect, 20 barrels of oil is equivalent to 40 tons of lumber: $20 \text{ oil} = 40 \text{ lumber}$. Divide both sides of the equation by 20 to calculate the opportunity cost of one barrel of oil in Canada. $20/20 \text{ oil} = 40/20 \text{ lumber}$. $1 \text{ oil} = 2 \text{ lumber}$. To produce one additional barrel of oil in Canada has an opportunity cost of 2 lumber. Calculate the same way for Venezuela: $60 \text{ oil} = 30 \text{ lumber}$. Divide both sides of the equation by 60. One oil in Venezuela has an opportunity cost of $1/2 \text{ lumber}$. Because $1/2 \text{ lumber} < 2 \text{ lumber}$, Venezuela has the comparative advantage in producing oil.

Step 4. Calculate the opportunity cost of one lumber by reversing the numbers, with lumber on the left side of the equation. In Canada, 40 lumber is equivalent in labor time to 20 barrels of oil: $40 \text{ lumber} = 20 \text{ oil}$. Divide each side of the equation by 40. The opportunity cost of one lumber is $1/2 \text{ oil}$. In Venezuela, the equivalent labor time will produce 30 lumber or 60 oil: $30 \text{ lumber} = 60 \text{ oil}$. Divide each side by 30. One lumber has an opportunity cost of two oil. Canada has the lower opportunity cost in producing lumber.

Step 5. In this example, absolute advantage is the same as comparative advantage. Canada has the absolute and comparative advantage in lumber; Venezuela has the absolute and comparative advantage in oil.

Step 6. Canada should specialize in what it has a relative lower opportunity cost, which is lumber, and Venezuela should specialize in oil. Canada will be exporting lumber and importing oil, and Venezuela will be exporting oil and importing lumber.

Comparative Advantage Goes Camping

To build an intuitive understanding of how comparative advantage can benefit all parties, set aside examples that involve national economies for a moment and consider the situation of a group of friends who decide to go camping together. The six friends have a wide range of skills and experiences, but one person in particular, Jethro, has done lots of camping before and is also a great athlete. Jethro has an absolute advantage in all aspects of camping: he is faster at carrying a backpack, gathering firewood, paddling a canoe, setting up tents, making a meal, and washing up. So here is the question: Because Jethro has an absolute productivity advantage in everything, should he do all the work?

Of course not! Even if Jethro is willing to work like a mule while everyone else sits around, he, like most mortals, only has 24 hours in a day. If everyone sits around and waits for Jethro to do everything, not only will Jethro be an unhappy camper, but there will not be much output for his group of six friends to consume. The theory of comparative advantage suggests that everyone will benefit if they figure out their areas of comparative advantage—that is, the area of camping where their productivity disadvantage is least, compared to Jethro. For example, it may be that Jethro is 80% faster at building fires and cooking meals than anyone else, but only 20% faster at gathering firewood and 10% faster at setting up tents. In that case, Jethro should focus on building fires and making meals, and others should attend to the other tasks, each according to where their productivity disadvantage is smallest. If the campers coordinate their efforts according to comparative advantage, they can all gain.

Key Concepts and Summary

Even when a country has high levels of productivity in all goods, it can still benefit from trade. Gains from trade come about as a result of comparative advantage. By specializing in a good that it gives up the least to produce, a country can produce more and offer that additional output for sale. If other countries specialize in the area of their comparative advantage as well and trade, the highly productive country is able to benefit from a lower opportunity cost of production in other countries.

Self-Check Question

Exercise:

Problem:

In Germany it takes three workers to make one television and four workers to make one video camera. In Poland it takes six workers to make one television and 12 workers to make one video camera.

- a. Who has the absolute advantage in the production of televisions? Who has the absolute advantage in the production of video cameras? How can you tell?
- b. Calculate the opportunity cost of producing one additional television set in Germany and in Poland. (Your calculation may involve fractions, which is fine.) Which country has a comparative advantage in the production of televisions?
- c. Calculate the opportunity cost of producing one video camera in Germany and in Poland. Which country has a comparative advantage in the production of video cameras?
- d. In this example, is absolute advantage the same as comparative advantage, or not?
- e. In what product should Germany specialize? In what product should Poland specialize?

Solution:

(a) In Germany, it takes fewer workers to make either a television or a video camera. Germany has an absolute advantage in the production of both goods.

(b) Producing an additional television in Germany requires three workers. Shifting those three German workers will reduce video camera production by $\frac{3}{4}$ of a camera. Producing an additional television set in Poland requires six workers, and shifting those workers from the other good reduces output of video cameras by $\frac{6}{12}$ of a camera, or $\frac{1}{2}$. Thus, the opportunity cost of producing televisions is lower in Poland, so Poland has the comparative advantage in the

production of televisions. *Note:* Do not let the fractions like $\frac{3}{4}$ of a camera or $\frac{1}{2}$ of a video camera bother you. If either country was to expand television production by a significant amount—that is, lots more than one unit—then we will be talking about whole cameras and not fractional ones. You can also spot this conclusion by noticing that Poland's absolute disadvantage is relatively lower in televisions, because Poland needs twice as many workers to produce a television but three times as many to produce a video camera, so the product with the relatively lower absolute disadvantage is Poland's comparative advantage.

(c) Producing a video camera in Germany requires four workers, and shifting those four workers away from television production has an opportunity cost of $\frac{4}{3}$ television sets. Producing a video camera in Poland requires 12 workers, and shifting those 12 workers away from television production has an opportunity cost of two television sets. Thus, the opportunity cost of producing video cameras is lower in Germany, and video cameras will be Germany's comparative advantage.

(d) In this example, absolute advantage differs from comparative advantage. Germany has the absolute advantage in the production of both goods, but Poland has a comparative advantage in the production of televisions. (e) Germany should specialize, at least to some extent, in the production of video cameras, export video cameras, and import televisions. Conversely, Poland should specialize, at least to some extent, in the production of televisions, export televisions, and import video cameras.

Review Questions

Exercise:

Problem:

Is it possible to have a comparative advantage in the production of a good but not to have an absolute advantage? Explain.

Exercise:

Problem: How does comparative advantage lead to gains from trade?

Critical Thinking Questions

Exercise:

Problem:

You just overheard your friend say the following: “Poor countries like Malawi have no absolute advantages. They have poor soil, low investments in formal education and hence low-skill workers, no capital, and no natural resources to speak of. Because they have no advantage, they cannot benefit from trade.” How would you respond?

Exercise:

Problem:

Look at [\[link\]](#). Is there a range of trades for which there will be no gains?

Exercise:

Problem:

You just got a job in Washington, D.C. You move into an apartment with some acquaintances. All your roommates, however, are slackers and do not clean up after themselves. You, on the other hand, can clean faster than each of them. You determine that you are 70% faster at dishes and 10% faster with vacuuming. All of these tasks have to be done daily. Which jobs should you assign to your roommates to get the most free time overall? Assume you have the same number of hours to devote to cleaning. Now, since you are faster, you seem to get done quicker than your roommate. What sorts of problems may this create? Can you imagine a trade-related analogy to this problem?

Problems

Exercise:**Problem:**

In Japan, one worker can make 5 tons of rubber or 80 radios. In Malaysia, one worker can make 10 tons of rubber or 40 radios.

- a. Who has the absolute advantage in the production of rubber or radios? How can you tell?
- b. Calculate the opportunity cost of producing 80 additional radios in Japan and in Malaysia. (Your calculation may involve fractions, which is fine.) Which country has a comparative advantage in the production of radios?
- c. Calculate the opportunity cost of producing 10 additional tons of rubber in Japan and in Malaysia. Which country has a comparative advantage in producing rubber?
- d. In this example, does each country have an absolute advantage and a comparative advantage in the same good?
- e. In what product should Japan specialize? In what product should Malaysia specialize?

Exercise:**Problem:**

Review the numbers for Canada and Venezuela from [\[link\]](#) which describes how many barrels of oil and tons of lumber the workers can produce. Use these numbers to answer the rest of this question.

- a. Draw a production possibilities frontier for each country. Assume there are 100 workers in each country. Canadians and Venezuelans desire both oil and lumber. Canadians want at least 2,000 tons of lumber. Mark a point on their production possibilities where they can get at least 3,000 tons.
- b. Assume that the Canadians specialize completely because they figured out they have a comparative advantage in lumber. They are willing to give up 1,000 tons of lumber. How much oil should they ask for in return for this lumber to be as well off as they were

with no trade? How much should they ask for if they want to gain from trading with Venezuela? *Note:* We can think of this “ask” as the relative price or trade price of lumber.

- c. Is the Canadian “ask” you identified in (b) also beneficial for Venezuelans? Use the production possibilities frontier graph for Venezuela to show that Venezuelans can gain from trade.

Exercise:

Problem:

In [\[link\]](#), is there an “ask” where Venezuelans may say “no thank you” to trading with Canada?

References

Bernstein, William J. *A Splendid Exchange: How Trade Shaped the World*. Atlantic Monthly Press. New York. 2008.

Intra-industry Trade between Similar Economies

By the end of this section, you will be able to:

- Identify at least two advantages of intra-industry trading
- Explain the relationship between economies of scale and intra-industry trade

Absolute and comparative advantages explain a great deal about patterns of global trade. For example, they help to explain the patterns noted at the start of this chapter, like why you may be eating fresh fruit from Chile or Mexico, or why lower productivity regions like Africa and Latin America are able to sell a substantial proportion of their exports to higher productivity regions like the European Union and North America. Comparative advantage, however, at least at first glance, does not seem especially well-suited to explain other common patterns of international trade.

The Prevalence of Intra-industry Trade between Similar Economies

The theory of comparative advantage suggests that trade should happen between economies with large differences in opportunity costs of production. Roughly half of all world trade involves shipping goods between the fairly similar high-income economies of the United States, Canada, the European Union, Japan, Mexico, and China (see [\[link\]](#)).

Country	U.S. Exports Go to ...	U.S. Imports Come from ...
European Union	19.0%	21.0%

Country	U.S. Exports Go to ...	U.S. Imports Come from ...
Canada	22.0%	14.0%
Japan	4.0%	6.0%
Mexico	15.0%	13.0%
China	8.0%	20.0%

Where U.S. Exports Go and U.S. Imports Originate (2015)(Source: https://www.census.gov/foreign-trade/Press-Release/current_press_release/ft900.pdf)

Moreover, the theory of comparative advantage suggests that each economy should specialize to a degree in certain products, and then exchange those products. A high proportion of trade, however, is **intra-industry trade**—that is, trade of goods within the same industry from one country to another. For example, the United States produces and exports autos and imports autos. [\[link\]](#) shows some of the largest categories of U.S. exports and imports. In all of these categories, the United States is both a substantial exporter and a substantial importer of goods from the same industry. In 2014, according to the Bureau of Economic Analysis, the United States exported \$159 billion worth of autos, and imported \$327 billion worth of autos. About 60% of U.S. trade and 60% of European trade is intra-industry trade.

Some U.S. Exports	Quantity of Exports (\$ billions)	Quantity of Imports (\$ billions)
--------------------------	--	--

Some U.S. Exports	Quantity of Exports (\$ billions)	Quantity of Imports (\$ billions)
Autos	\$146	\$327
Food and beverages	\$144	\$126
Capital goods	\$550	\$551
Consumer goods	\$199	\$558
Industrial supplies	\$507	\$665
Other transportation	\$45	\$55

Some Intra-Industry U.S. Exports and Imports in 2014(Source: <http://www.bea.gov/newsreleases/international/trade/tradnewsrelease.htm>)

Why do similar high-income economies engage in intra-industry trade? What can be the economic benefit of having workers of fairly similar skills making cars, computers, machinery and other products which are then shipped across the oceans to and from the United States, the European Union, and Japan? There are two reasons: (1) The division of labor leads to learning, innovation, and unique skills; and (2) economies of scale.

Gains from Specialization and Learning

Consider the category of machinery, where the U.S. economy has considerable intra-industry trade. Machinery comes in many varieties, so the United States may be exporting machinery for manufacturing with wood, but importing machinery for photographic processing. The underlying reason why a country like the United States, Japan, or Germany

produces one kind of machinery rather than another is usually not related to U.S., German, or Japanese firms and workers having generally higher or lower skills. It is just that, in working on very specific and particular products, firms in certain countries develop unique and different skills.

Specialization in the world economy can be very finely split. In fact, recent years have seen a trend in international trade called **splitting up the value chain**. The **value chain** describes how a good is produced in stages. As indicated in the beginning of the chapter, the production of the iPhone involves the design and engineering of the phone in the United States, parts supplied from Korea, the assembly of the parts in China, and the advertising and marketing done in the United States. Thanks in large part to improvements in communication technology, sharing information, and transportation, it has become easier to split up the value chain. Instead of production in a single large factory, all of these steps can be split up among different firms operating in different places and even different countries. Because firms split up the value chain, international trade often does not involve whole finished products like automobiles or refrigerators being traded between nations. Instead, it involves shipping more specialized goods like, say, automobile dashboards or the shelving that fits inside refrigerators. Intra-industry trade between similar countries produces economic gains because it allows workers and firms to learn and innovate on particular products—and often to focus on very particular parts of the value chain.

Note:

Visit this [website](#) for some interesting information about the assembly of the iPhone.

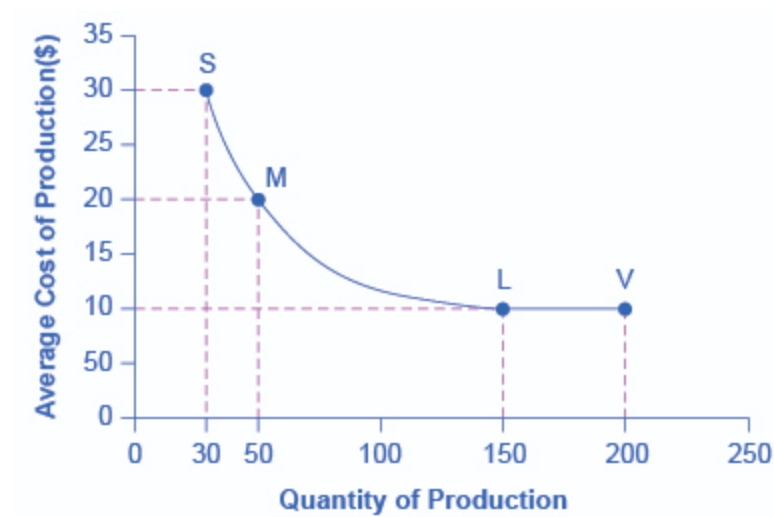


Economies of Scale, Competition, Variety

A second broad reason that intra-industry trade between similar nations produces economic gains involves economies of scale. The concept of economies of scale, as introduced in [Cost and Industry Structure](#), means that as the scale of output goes up, average costs of production decline—at least up to a point. [\[link\]](#) illustrates economies of scale for a plant producing toaster ovens. The horizontal axis of the figure shows the quantity of production by a certain firm or at a certain manufacturing plant. The vertical axis measures the average cost of production. Production plant S produces a small level of output at 30 units and has an average cost of production of \$30 per toaster oven. Plant M produces at a medium level of output at 50 units, and has an average cost of production of \$20 per toaster oven. Plant L produces 150 units of output with an average cost of production of only \$10 per toaster oven. Although plant V can produce 200 units of output, it still has the same unit cost as Plant L.

In this example, a small or medium plant, like S or M, will not be able to compete in the market with a large or a very large plant like L or V, because the firm that operates L or V will be able to produce and sell their output at a lower price. In this example, economies of scale operate up to point L, but beyond point L to V, the additional scale of production does not continue to reduce average costs of production.

Economies of Scale



Production Plant S, has an average cost of

production of \$30 per toaster oven.
Production plant M has an average cost of
production of \$20 per toaster oven.
Production plant L has an average cost of
production of only \$10 per toaster oven.
Production plant V would still have an
average cost of production of \$10 per
toaster oven. Thus, production plant M can
produce toaster ovens more cheaply than
plant S because of economies of scale, and
plants L or V can produce more cheaply
than S or M because of economies of scale.
However, the economies of scale end at an
output level of 150. Plant V, despite being
larger, cannot produce more cheaply on
average than plant L.

The concept of economies of scale becomes especially relevant to international trade when it enables one or two large producers to supply the entire country. For example, a single large automobile factory could probably supply all the cars purchased in a smaller economy like the United Kingdom or Belgium in a given year. However, if a country has only one or two large factories producing cars, and no international trade, then consumers in that country would have relatively little choice between kinds of cars (other than the color of the paint and other nonessential options). Little or no competition will exist between different car manufacturers.

International trade provides a way to combine the lower average production costs that come from economies of scale and still have competition and variety for consumers. Large automobile factories in different countries can make and sell their products around the world. If the U.S. automobile market was made up of only General Motors, Ford, and Chrysler, the level of competition and consumer choice would be quite a lot lower than when U.S. carmakers must face competition from Toyota, Honda, Suzuki, Fiat, Mitsubishi, Nissan, Volkswagen, Kia, Hyundai, BMW, Subaru, and others.

Greater competition brings with it innovation and responsiveness to what consumers want. America's car producers make far better cars now than they did several decades ago, and much of the reason is competitive pressure, especially from East Asian and European carmakers.

Dynamic Comparative Advantage

The sources of gains from intra-industry trade between similar economies—namely, the learning that comes from a high degree of specialization and splitting up the value chain and from economies of scale—do not contradict the earlier theory of comparative advantage. Instead, they help to broaden the concept.

In intra-industry trade, the level of worker productivity is not determined by climate or geography. It is not even determined by the general level of education or skill. Instead, the level of worker productivity is determined by how firms engage in specific learning about specialized products, including taking advantage of economies of scale. In this vision, comparative advantage can be dynamic—that is, it can evolve and change over time as new skills are developed and as the value chain is split up in new ways. This line of thinking also suggests that countries are not destined to have the same comparative advantage forever, but must instead be flexible in response to ongoing changes in comparative advantage.

Key Concepts and Summary

A large share of global trade happens between high-income economies that are quite similar in having well-educated workers and advanced technology. These countries practice intra-industry trade, in which they import and export the same products at the same time, like cars, machinery, and computers. In the case of intra-industry trade between economies with similar income levels, the gains from trade come from specialized learning in very particular tasks and from economies of scale. Splitting up the value chain means that several stages of producing a good take place in different countries around the world.

Self-Check Questions

Exercise:

Problem:

How can there be any economic gains for a country from both importing and exporting the same good, like cars?

Solution:

There are a number of possible advantages of intra-industry trade. Both nations can take advantage of extreme specialization and learning in certain kinds of cars with certain traits, like gas-efficient cars, luxury cars, sport-utility vehicles, higher- and lower-quality cars, and so on. Moreover, nations can take advantage of economies of scale, so that large companies will compete against each other across international borders, providing the benefits of competition and variety to customers. This same argument applies to trade between U.S. states, where people often buy products made by people of other states, even though a similar product is made within the boundaries of their own state. All states—and all countries—can benefit from this kind of competition and trade.

Exercise:

Problem:

[\[link\]](#) shows how the average costs of production for semiconductors (the “chips” in computer memories) change as the quantity of semiconductors built at that factory increases.

- a. Based on these data, sketch a curve with quantity produced on the horizontal axis and average cost of production on the vertical axis. How does the curve illustrate economies of scale?
- b. If the equilibrium quantity of semiconductors demanded is 90,000, can this economy take full advantage of economies of scale? What about if quantity demanded is 70,000?

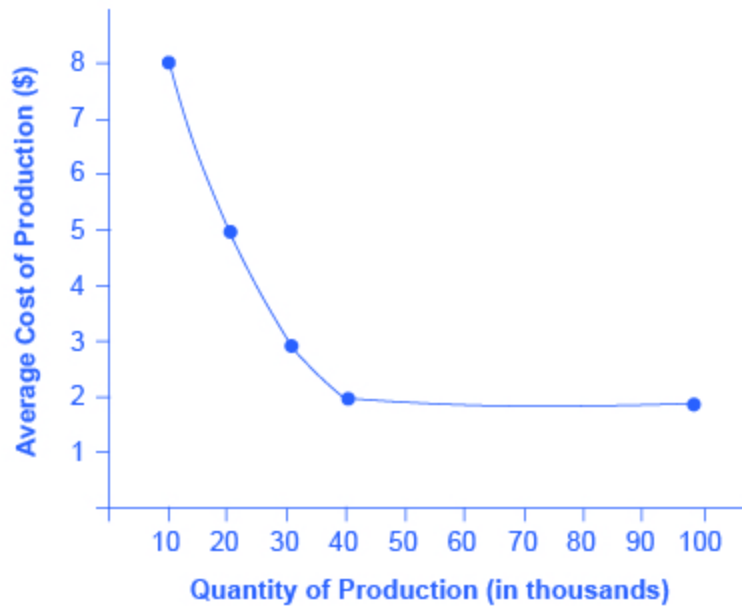
semiconductors? 50,000 semiconductors? 30,000 semiconductors?

- c. Explain how international trade could make it possible for even a small economy to take full advantage of economies of scale, while also benefiting from competition and the variety offered by several producers.

Quantity of Semiconductors	Average Total Cost
10,000	\$8 each
20,000	\$5 each
30,000	\$3 each
40,000	\$2 each
100,000	\$2 each

Solution:

(a) Start by plotting the points on a sketch diagram and then drawing a line through them. The following figure illustrates the average costs of production of semiconductors.



The curve illustrates economies of scale by showing that as the scale increases—that is, as production at this particular factory goes up—the average cost of production declines. The economies of scale exist up to an output of 40,000 semiconductors; at higher outputs, the average cost of production does not seem to decline any further.

(b) At any quantity demanded above 40,000, this economy can take full advantage of economies of scale; that is, it can produce at the lowest cost per unit. Indeed, if the quantity demanded was quite high, like 500,000, then there could be a number of different factories all taking full advantage of economies of scale and competing with each other. If the quantity demanded falls below 40,000, then the economy by itself, without foreign trade, cannot take full advantage of economies of scale.

(c) The simplest answer to this question is that the small country could have a large enough factory to take full advantage of economies of scale, but then export most of the output. For semiconductors, countries like Taiwan and Korea have recently fit this description. Moreover, this country could also import semiconductors from other countries which also have large factories, thus getting the benefits of

competition and variety. A slightly more complex answer is that the country can get these benefits of economies of scale without producing semiconductors, but simply by buying semiconductors made at low cost around the world. An economy, especially a smaller country, may well end up specializing and producing a few items on a large scale, but then trading those items for other items produced on a large scale, and thus gaining the benefits of economies of scale by trade, as well as by direct production.

Review Questions

Exercise:

Problem: What is intra-industry trade?

Exercise:

Problem:

What are the two main sources of economic gains from intra-industry trade?

Exercise:

Problem: What is splitting up the value chain?

Critical Thinking Questions

Exercise:

Problem:

Does intra-industry trade contradict the theory of comparative advantage?

Exercise:

Problem: Do consumers benefit from intra-industry trade?

Exercise:**Problem:**

Why might intra-industry trade seem surprising from the point of view of comparative advantage?

Problems**Exercise:****Problem:**

From earlier chapters you will recall that technological change shifts the average cost curves. Draw a graph showing how technological change could influence intra-industry trade.

Exercise:**Problem:**

Consider two countries: South Korea and Taiwan. Taiwan can produce one million mobile phones per day at the cost of \$10 per phone and South Korea can produce 50 million mobile phones at \$5 per phone. Assume these phones are the same type and quality and there is only one price. What is the minimum price at which both countries will engage in trade?

References

U.S. Census Bureau. 2015. "U.S. International Trade in Goods and Services: December 2014." Accessed April 13, 2015.
<http://www.bea.gov/newsreleases/international/trade/2015/pdf/trad1214.pdf>
.

U.S. Census Bureau. U.S. Bureau of Economic Analysis. 2015. "U.S. International Trade in Goods and Services February 2015." Accessed April

10, 2015. https://www.census.gov/foreign-trade/Press-Release/current_press_release/ft900.pdf.

Vernengo, Matias. “What Do Undergraduates Really Need to Know About Trade and Finance?” in *Political Economy and Contemporary Capitalism: Radical Perspectives on Economic Theory and Policy*, ed. Ron Baiman, Heather Boushey, and Dawn Saunders. M. E. Sharpe Inc, 2000. Armonk. 177-183.

Glossary

intra-industry trade

international trade of goods within the same industry

splitting up the value chain

many of the different stages of producing a good happen in different geographic locations

value chain

how a good is produced in stages

The Benefits of Reducing Barriers to International Trade

By the end of this section, you will be able to:

- Explain tariffs as barriers to trade
- Identify at least two benefits of reducing barriers to international trade

Tariffs are taxes that governments place on imported goods for a variety of reasons. Some of these reasons include protecting sensitive industries, for humanitarian reasons, and protecting against dumping. Traditionally, tariffs were used simply as a political tool to protect certain vested economic, social, and cultural interests. The World Trade Organization (WTO) is committed to lowering barriers to trade. The world's nations meet through the WTO to negotiate how they can reduce barriers to trade, such as tariffs. WTO negotiations happen in "rounds," where all countries negotiate one agreement to encourage trade, take a year or two off, and then start negotiating a new agreement. The current round of negotiations is called the Doha Round because it was officially launched in Doha, the capital city of Qatar, in November 2001. In 2009, economists from the World Bank summarized recent research and found that the Doha round of negotiations would increase the size of the world economy by \$160 billion to \$385 billion per year, depending on the precise deal that ended up being negotiated.

In the context of a global economy that currently produces more than \$30 trillion of goods and services each year, this amount is not huge: it is an increase of 1% or less. But before dismissing the gains from trade too quickly, it is worth remembering two points.

- First, a gain of a few hundred billion dollars is enough money to deserve attention! Moreover, remember that this increase is not a one-time event; it would persist each year into the future.
- Second, the estimate of gains may be on the low side because some of the gains from trade are not measured especially well in economic statistics. For example, it is difficult to measure the potential advantages to consumers of having a variety of products available and a greater degree of competition among producers. Perhaps the most important unmeasured factor is that trade between countries, especially

when firms are splitting up the value chain of production, often involves a transfer of knowledge that can involve skills in production, technology, management, finance, and law.

Low-income countries benefit more from trade than high-income countries do. In some ways, the giant U.S. economy has less need for international trade, because it can already take advantage of internal trade within its economy. However, many smaller national economies around the world, in regions like Latin America, Africa, the Middle East, and Asia, have much more limited possibilities for trade inside their countries or their immediate regions. Without international trade, they may have little ability to benefit from comparative advantage, slicing up the value chain, or economies of scale. Moreover, smaller economies often have fewer competitive firms making goods within their economy, and thus firms have less pressure from other firms to provide the goods and prices that consumers want.

The economic gains from expanding international trade are measured in hundreds of billions of dollars, and the gains from international trade as a whole probably reach well into the trillions of dollars. The potential for gains from trade may be especially high among the smaller and lower-income countries of the world.

Note:

Visit this [website](#) for a list of some benefits of trade.



From Interpersonal to International Trade

Most people find it easy to believe that they, personally, would not be better off if they tried to grow and process all of their own food, to make all of their own clothes, to build their own cars and houses from scratch, and so on. Instead, we all benefit from living in economies where people and firms can specialize and trade with each other.

The benefits of trade do not stop at national boundaries, either. Earlier we explained that the division of labor could increase output for three reasons: (1) workers with different characteristics can specialize in the types of production where they have a comparative advantage; (2) firms and workers who specialize in a certain product become more productive with learning and practice; and (3) economies of scale. These three reasons apply from the individual and community level right up to the international level. If it makes sense to you that interpersonal, intercommunity, and interstate trade offer economic gains, it should make sense that international trade offers gains, too.

International trade currently involves about \$20 trillion worth of goods and services moving around the globe. Any economic force of that size, even if it confers overall benefits, is certain to cause disruption and controversy. This chapter has only made the case that trade brings economic benefits. Other chapters discuss, in detail, the public policy arguments over whether to restrict international trade.

Note:

It's Apple's (Global) iPhone

Apple Corporation uses a global platform to produce the iPhone. Now that you understand the concept of comparative advantage, you can see why the engineering and design of the iPhone is done in the United States. The United States has built up a comparative advantage over the years in designing and marketing products, and sacrifices fewer resources to design high-tech devices relative to other countries. China has a comparative advantage in assembling the phone due to its large skilled labor force. Korea has a comparative advantage in producing components. Korea focuses its production by increasing its scale, learning better ways to produce screens and computer chips, and uses innovation to lower average

costs of production. Apple, in turn, benefits because it can purchase these quality products at lower prices. Put the global assembly line together and you have the device with which we are all so familiar.

Key Concepts and Summary

Tariffs are placed on imported goods as a way of protecting sensitive industries, for humanitarian reasons, and for protection against dumping. Traditionally, tariffs were used as a political tool to protect certain vested economic, social, and cultural interests. The WTO has been, and continues to be, a way for nations to meet and negotiate through barriers to trade. The gains of international trade are very large, especially for smaller countries, but are beneficial to all.

Self-Check Question

Exercise:

Problem:

If the removal of trade barriers is so beneficial to international economic growth, why would a nation continue to restrict trade on some imported or exported products?

Solution:

A nation might restrict trade on imported products to protect an industry that is important for national security. For example, nation X and nation Y may be geopolitical rivals, each with ambitions of increased political and economic strength. Even if nation Y has comparative advantage in the production of missile defense systems, it is unlikely that nation Y would seek to export those goods to nation X. It is also the case that, for some nations, the production of a particular good is a key component of national identity. In Japan, the production of rice is culturally very important. It may be difficult for Japan to

import rice from a nation like Vietnam, even if Vietnam has a comparative advantage in rice production.

Review Question

Exercise:

Problem:

Are the gains from international trade more likely to be relatively more important to large or small countries?

Critical Thinking Questions

Exercise:

Problem:

In World Trade Organization meetings, what do you think low-income countries lobby for?

Exercise:

Problem:

Why might a low-income country put up barriers to trade, such as tariffs on imports?

Exercise:

Problem:

Can a nation's comparative advantage change over time? What factors would make it change?

Problems

Exercise:

Problem:

If trade increases world GDP by 1% per year, what is the global impact of this increase over 10 years? How does this increase compare to the annual GDP of a country like Sri Lanka? Discuss. *Hint:* To answer this question, here are steps you may want to consider. Go to the World Development Indicators (online) published by the World Bank. Find the current level of World GDP in constant international dollars. Also, find the GDP of Sri Lanka in constant international dollars. Once you have these two numbers, compute the amount the additional increase in global incomes due to trade and compare that number to Sri Lanka's GDP.

References

World Trade Organization. "The Doha Round." Accessed October 2013. http://www.wto.org/english/tratop_e/dda_e/dda_e.htm.

The World Bank. "Data: World Development Indicators." Accessed October 2013. <http://data.worldbank.org/data-catalog/world-development-indicators>.

Glossary

tariffs

taxes that governments place on imported goods